

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 14, 2000, 18:45:20 ; Search time 871.67 Seconds
(without alignments)
5114.947 Million cell updates/sec

Title: US-09-223-796-1
Perfect score: 1100
Sequence: 1 gcacgagccacagccagcta.....attaaaaaaaaaaaaa 1100

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 4857316 seqs, 2026611650 residues

Total number of hits satisfying chosen parameters: 9714632

Minimum DB seq length: 0
Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database :

- EST.*
- 1: em_est1.*
 - 2: em_est2.*
 - 3: em_est3.*
 - 4: em_est4.*
 - 5: em_est5.*
 - 6: em_est6.*
 - 7: em_est7.*
 - 8: em_est8.*
 - 9: em_est9.*
 - 10: em_est10.*
 - 11: em_est11.*
 - 12: em_est12.*
 - 13: em_est13.*
 - 14: em_est14.*
 - 15: em_est15.*
 - 16: em_est16.*
 - 17: em_est17.*
 - 18: em_est18.*
 - 19: em_est19.*
 - 20: gb_est1.*
 - 21: gb_est2.*
 - 22: gb_est3.*
 - 23: gb_est4.*
 - 24: gb_est5.*
 - 25: gb_est6.*
 - 26: gb_est7.*
 - 27: gb_est8.*
 - 28: gb_est9.*
 - 29: gb_est10.*
 - 30: gb_est11.*
 - 31: gb_est12.*
 - 32: gb_est13.*
 - 33: gb_est14.*
 - 34: gb_est15.*
 - 35: gb_est16.*
 - 36: gb_est17.*
 - 37: gb_est18.*
 - 38: gb_est19.*
 - 39: gb_est20.*
 - 40: gb_est21.*
 - 41: gb_est22.*
 - 42: gb_est23.*
 - 43: gb_est24.*
 - 44: gb_est25.*

- 45: gb_est26.*
- 46: gb_est27.*
- 47: gb_est28.*
- 48: gb_est29.*
- 49: gb_est30.*
- 50: gb_est31.*
- 51: gb_est32.*
- 52: em_est20.*
- 53: em_est21.*
- 54: em_est22.*
- 55: em_est23.*
- 56: em_est24.*
- 57: em_est25.*
- 58: em_est26.*
- 59: gb_est33.*
- 60: gb_est34.*
- 61: gb_est35.*
- 62: gb_est36.*
- 63: gb_est37.*
- 64: gb_est38.*
- 65: em_est27.*
- 66: em_est28.*
- 67: em_est29.*
- 68: em_est30.*
- 69: gb_est39.*
- 70: gb_est40.*
- 71: gb_est41.*
- 72: gb_est42.*
- 73: gb_est43.*
- 74: gb_est44.*
- 75: em_est31.*
- 76: em_est32.*
- 77: em_est33.*
- 78: em_est34.*
- 79: gb_est45.*
- 80: gb_est46.*
- 81: gb_gss1.*
- 82: gb_gss2.*
- 83: gb_gss3.*
- 84: gb_gss4.*
- 85: gb_gss5.*
- 86: em_gss1.*
- 87: em_gss2.*
- 88: em_gss3.*
- 89: em_gss4.*
- 90: gb_gss5.*
- 91: gb_gss6.*
- 92: gb_gss7.*
- 93: gb_gss8.*
- 94: gb_gss9.*
- 95: em_gss5.*
- 96: em_gss6.*
- 97: em_gss7.*
- 98: em_gss8.*
- 99: em_gss9.*
- 100: em_gss10.*
- 101: em_gss11.*
- 102: gb_gss10.*
- 103: gb_gss11.*
- 104: em_gss12.*
- 105: gb_gss12.*
- 106: gb_gss13.*
- 107: gb_gss14.*
- 108: gb_gss15.*
- 109: gb_gss16.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result Query

No.	Score	Match	Length	DB	ID	Description
1	595.6	54.1	602	43	AI104547	AI104547 EST213836
2	570.4	51.9	680	45	AI232721	AI232721 EST229409
3	446.6	40.6	462	43	AI233681	AI233681 EST230369
4	383.4	34.9	486	27	AA034568	AA034568 m149a11.r
5	382.8	34.8	456	42	AI136965	AI136965 UI-R-C2p
6	376.6	34.2	482	41	AI048873	AI048873 uc76f02.y
7	373	33.9	444	27	AA032667	AA032667 m132b07.y
8	362	32.9	454	42	AI154070	AI154070 u476a06.r
9	334.4	30.4	345	81	AA434638	AA434638 UI-R-EJOP
10	327.4	29.8	540	94	AA035548	AA035548 CIT-RSP-2
11	307.4	27.9	394	28	AA065442	AA065442 m138h02.r
12	302.4	27.5	446	27	W83837	W83837 mf26c06.r1
13	296.6	27.0	569	36	C77953	C77953 C77953 Mous
14	296.4	26.9	374	33	AA450933	AA450933 v185a11.r
15	286.6	26.1	574	36	C78711	C78711 C78711 Mous
16	285	25.9	350	26	W67032	W67032 me29f10.r1
17	276	25.1	501	90	AQ798652	AQ798652 HS_2238_B
18	260.2	23.7	537	27	AA029831	AA029831 z1k10b06.r
19	243.4	22.1	548	38	AA780068	AA780068 z124h04.s
20	234.4	21.3	260	30	AA239639	AA239639 mv25a05.r
21	234.2	21.3	487	45	AI178891	AI178891 EST222573
22	233	21.2	300	28	AA067980	AA067980 mm56f09.r
23	224.2	20.4	520	63	AA005581	AA005581 w287h11.x
24	214.4	19.5	337	81	AA418433	AA418433 56258 MAR
25	213.8	19.4	525	61	AI871472	AI871472 w167e01.x
26	213.8	19.4	527	64	AA025527	AA025527 wu97d10.x
27	204.6	18.6	424	44	AU045080	AU045080 AU045080
28	195.2	17.7	569	44	AU043460	AU043460 AU043460
29	190.6	17.3	464	26	W86176	W86176 zh59g11.r1
30	177	16.1	318	48	AI578514	AI578514 UI-R-AA0-
31	174.4	15.9	491	26	W93417	W93417 zd95d01.s1
32	171.6	15.6	447	39	AA836105	AA836105 od15c01.s
33	164.6	15.0	431	43	AI240879	AI240879 q194e02.x
34	164.6	15.0	437	44	AI276262	AI276262 q165f06.x
35	158.6	14.4	413	42	AI076231	AI076231 oy65c09.x
36	157.6	14.3	411	79	AA261921	AA261921 x87f11.x
37	157.6	14.3	418	43	AI189612	AI189612 q329g06.x
38	155.6	14.1	330	32	AA379908	AA379908 EST92801
39	154.4	14.0	361	26	W92672	W92672 zh79g09.r1
40	152.6	13.9	406	34	AA464962	AA464962 z880f06.s
41	150.2	13.7	410	39	AA857959	AA857959 o639g04.s
42	148.4	13.5	398	43	AI200129	AI200129 qf90h05.x
43	148.4	13.5	466	21	R20463	R20463 yg33h09.r1
44	146.6	13.3	418	26	W86177	W86177 zh59g11.s1
45	145.8	13.3	447	45	AI376292	AI376292 ta61c07.x

ALIGNMENTS

RESULT 1	AI104547	602 bp	mRNA	EST	31-JAN-1999
LOCUS	AI104547/c				
DEFINITION	EST213836	Normalized rat heart, Bento Soares	Rattus sp.	cdna clone	
ACCESSION	RHECE94	3' end, mRNA sequence.			
VERSION	AI104547				
KEYWORDS	EST.	AI104547.1	GI:3708888		
SOURCE	Rattus sp.				
ORGANISM	Eukaryota; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.				
REFERENCE	Lee,N.H., Glodek,A., Chandra,I., Mason,T.M., Quackenbush,J., Kerlavage,A.R. and Adams,M.D.				
AUTHORS	Lee,N.H., Glodek,A., Chandra,I., Mason,T.M., Quackenbush,J., Kerlavage,A.R. and Adams,M.D.				
TITLE	Rat Genome Project: Generation of a Rat EST (RST) Catalog & Rat Gene Index				
JOURNAL	Unpublished (1998)				
COMMENT	Other ESTs: TC54601				
	Contact: Lee, NH				
	ATCC				

The Institute for Genomic Research
9712, Medical Center Drive, Rockville, MD 20850, USA
Tel: (301)-838-3529
Fax: (301)-838-0208
Email: shlee@tigr.org
Seq primer: M13-21.
Location/Qualifiers
source
1.602
/organism="Rattus sp."
/db_xref="ATCC (inhost):2025533"
/db_xref="taxon:10118"
/clone="RHECE94"
/clone_lib="Normalized rat heart, Bento Soares"
/note="Organ: heart; Vector: pT7r3Pac; Site_1: EcoRI; Site_2: NotI"
BASE COUNT 144 a 171 c 145 g 142 t
ORIGIN

Query Match 54.1%; Score 595.6; DB 43; Length 602;
Best Local Similarity 99.3%; Pred. No. 3.3e-157;
Matches 598; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy	482	cattctcagagatcaattggagatttggccagtttggcatttgggagtgcaagccctct	541
Db	602	CATTCTCAGGATCAATTGGAGATTGGCCAGTTTGGCAATTTGGAAGTCAACGCCCTCT	543
Qy	542	tctcgactctgtagcccaacagcagagatctcgctcctcagtgcttacttccggtg	601
Db	542	TCTCGACTCTGTAGCCCAACAGCAGGATCTCTGCTGCCACAGTGTCTTACTTCCGGTG	483
Qy	602	gcgggtgagctggccatctcaaccagcgtcagtcctccctgcaacccgagtggtct	661
Db	482	CCGGGTGGAGCTGGCCATCTCAACACCGCTCAGTCCCGCTCCCTGACCGAGTGTCT	423
Qy	662	catcgactgaagctcagatgattgacacccgttcgaggtgcccataagcaatt	721
Db	422	CATCGAGCTGAAGCTCAGATGATGATCGACACCGCTTCGAGGTGCCCAAAAT	363
Qy	722	tcaggagctcggtacagtgtagcttggcttcaagagatgacagactggagaaga	781
Db	362	TCAGGAGCTGGCGTACAGTGTAGCTTGGCTTAAAGAGATGCCAAGTGGAGAAGAA	303
Qy	782	gtgtgagcgcaaaactgcagagactgaactgaacccctggtactgtggtgctgaagctgtac	841
Db	302	GTGTGAGCGCAAACTGCAGGACTGACTGAACCCCTGGTACTGTGGGTGCTGAAGCTGTAC	243
Qy	842	cagaacacagccccccactgggtgatgagcccaactccattgaggtcctgcatgtgagaac	901
Db	242	CAGAACACAGCCCCCCTGGTGTATGAGGCCAACTCCATTGAGGTCTGTCATGTGAGAAC	183
Qy	902	gtatttlaagtgaagacagcggagcttccaggttttctttatgagtcacacagctggg	961
Db	182	GTAATTTAAGTGAAGACACAGCGGACTTTTCAGGTTTTTGTATGAGTCAACAGCTGGG	123
Qy	962	cagggtgacacagtttataatctcagcccttggagctgtgagctgaggaatgggaagt	1021
Db	122	CAGGTTGGCACAGTTTATAATCTCAGCCCTTGGAAAGCTGAGGCTGGGAATGGGAAGTG	63
Qy	1022	taagctgggccccttcttcagagtgaggtcagtgatggaattaaagaggtaaaagcaacta	1081
Db	62	TAAGCTGGGCCCTGGCTTTCATAGTGAGGCTCAGTGTCAATTAAGAGAGGTAAGCAACTA	3
Qy	1082	tt 1083	
Db	2	Tt 1	

RESULT 2
AI232721/c 680 bp mRNA EST 08-JAN-1999
LOCUS
DEFINITION EST229409 Normalized rat kidney, Bento Soares Rattus sp. cdna clone
RKICF61 3' end, mRNA sequence.

FEATURES	seq primer	Location/Qualifiers
source	1..462	/organism="Rattus sp." /db_xref="taxon:10118" /clone="RKIDI81" /clone_lib="Normalized rat kidney, Bento Soares" /note="Organ: kidney; Vector: pT73Pac; Site_1: EcoRI; Site_2: NotI"
BASE COUNT	108 a	133 c 100 g 121 t
ORIGIN		
Query Match	40.6%	Score 446.6; DB 43; Length 462;
Best Local Similarity	99.1%	Pred. No. 2.9e-113;
Matches 449; Conservative	0;	Mismatches 4; Indels 0; Gaps 0
Qy 638	cgcgtccctcgaaccagtggttctcatcgactgagctcacagatgatgctgcacacg	697
Db 462	CCGCTGCCCTGCACACCAGTGTTCTCATCGAGCTGAAGTCACAGATGGATGTGCACACCG	403
Qy 698	cttcgaggtgccatagcagcaatttcaggagctgcggtagcttagccttgtgccttaa	757
Db 402	CYTGCAGGTGCCATAGCCAAACTTCAGSAGCTGCGGTACAGTGTAGCCTTGGTCCTTAA	343
Qy 758	ggagtggcagaactggagaaagaagtgtgagcgcaaatcgaggactgactgaacctgg	817
Db 342	GGAGATGGCAGAACCTGGAGAAAGTGTGAGGCGCAACTGCAGGACTGACTGAACCCCTGG	283
Qy 818	tacttgagtgctgaagctggtaccagaacacagccccccactggtgatgagcccacatc	877
Db 282	TACTTGGGTGCTGAAGCTGGTACCAGAACACAGCCCCCCTGTTGTTATGAGCCCAATC	223
Qy 878	cattdaggtccctgcattgtagaacgtattttaagtgaaaagacagcgggagctttcaggtt	937
Db 222	CATTGAGGTCCCTGCATGTGCAGAACGTATTTTAAGTGAAAAGACAGCGGGACITTCAGGTT	163
Qy 938	ttgttttatgattcaacagctgggcagggtggcacagtttataatctcagcccttggaag	997
Db 162	TTGTTTTTATGAGTCAACAGCTGGCAGGTGGCACAGTTTATAATCTCAGCCCTTGGGAG	103

QY 998 tctgagcgtggaagaatgaagctgtaagctggcctggcttttcattatagtgagcgtcagtg 1057
 |||||
 Db 102 GCTGAGGCTGGAGAAATGGGAAGTGTAAAGCTGGGCTGGCTTTTCATAGTCAGGCTCAGTCT 43
 |||||
 QY 1058 cgaattaaagaggttaagcaactattaaaaaa 1090
 |||||
 Db 42 CGAATTAAGAGGCTAAAGCAACTATTATAAACA 10
 |||||

RESULT 4
 AA034568 456 bp mRNA EST 23-AUG-1996
 LOCUS mi43all.r1 Soares mouse embryo NBME13.5 14.5 Mus musculus cDNA
 DEFINITION clone IMAGE:466844 5', mRNA sequence.
 ACCESSION AA034568
 VERSION AA034568.1 GI:1506431
 KEYWORDS EST.
 SOURCE house mouse.
 ORGANISM Mus musculus.
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 REFERENCE 1 (bases 1 to 456)
 AUTHORS Marra, M., Hillier, L., Allen, M., Bowles, M., Dietrich, N., Dubuque, T.,
 Geisel, S., Kucaba, T., Lacy, M., Le, M., Martin, J., Morris, M.,
 Schellenberg, K., Steptoe, M., Tan, F., Underwood, K., Moore, B.,
 Theising, B., Wylie, T., Lennon, G., Soares, B., Wilson, R. and
 Waterston, R.
 TITLE The WashU-HHMI Mouse EST Project
 JOURNAL Unpublished (1996)
 COMMENT On Nov 29, 1993 this sequence version replaced gi:634431.
 Contact: Marra M/Mouse EST Project
 WashU-HHMI Mouse EST Project
 Washington University School of Medicine
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
 Tel: 314 286 1800
 Fax: 314 286 1810
 Email: mouseest@watson.wustl.edu
 This clone is available royalty-free through LNL ; contact the
 IMAGE Consortium (info@image.llnl.gov) for further information.
 MGI:280660
 Seq primer: -28M13 rev2 from Amer sham
 High quality sequence stop: 421.
 Location/Qualifiers
 1. 456
 /organism="Mus musculus"
 /strain="C57BL/6J"
 /db_xref="taxon:10090"
 /clone="IMAGE:466844"
 /clone_lib="Soares mouse embryo NBME13.5 14.5"
 /sex="unknown"
 /tissue_type="embryo"
 /dev_stage="13.5-14.5dpc total fetus"
 /lab_host="DH10B"
 /note="Vector: p7T73D-Pac (Pharmacia) with a modified
 polylinker. Site 1: Not I; Site 2: Eco RI; 1st strand cDNA
 was primed with a Not I - oligo(dT) primer [5'
 TGTTCACCACTGAAGTGGGAGCGCGCGGAAATTTTTTTTTTTTTTTTTTTT
 T 3'], on equal amounts of mRNA from 2 13.5dpc and 2
 14.5dpc embryos (total RNA provided by Minoru Ko, Wayne
 State Univ., from 2]; double-stranded cDNA was ligated to
 Eco RI adaptors (Pharmacia), digested with Not I and
 cloned into the Not I and Eco RI sites of the modified
 pT73 vector. Library went through one round of
 normalization, and was constructed by Bento Soares and
 M. Fatima Bonaldo."

FEATURES

Source
 1. 456
 /organism="Mus musculus"
 /strain="C57BL/6J"
 /db_xref="taxon:10090"
 /clone="IMAGE:466844"
 /clone_lib="Soares mouse embryo NBME13.5 14.5"
 /sex="unknown"
 /tissue_type="embryo"
 /dev_stage="13.5-14.5dpc total fetus"
 /lab_host="DH10B"
 /note="Vector: p7T73D-Pac (Pharmacia) with a modified
 polylinker. Site 1: Not I; Site 2: Eco RI; 1st strand cDNA
 was primed with a Not I - oligo(dT) primer [5'
 TGTTCACCACTGAAGTGGGAGCGCGCGGAAATTTTTTTTTTTTTTTTTTTT
 T 3'], on equal amounts of mRNA from 2 13.5dpc and 2
 14.5dpc embryos (total RNA provided by Minoru Ko, Wayne
 State Univ., from 2]; double-stranded cDNA was ligated to
 Eco RI adaptors (Pharmacia), digested with Not I and
 cloned into the Not I and Eco RI sites of the modified
 pT73 vector. Library went through one round of
 normalization, and was constructed by Bento Soares and
 M. Fatima Bonaldo."
 97 a 134 c 131 g 94 t

BASE COUNT

Query Match 34.98; Score 383.4; DB 27; Length 456;
 Best Local Similarity 91.88; Pred. No. 1.8e-97;
 Matches 405; Conservative 0; Mismatches 36; Indels 0; Gaps 0;

QY 77 atgcaactgtgttaggaccttcttctcgagctcgaagaacggaagaagatgtc 136
 |||||
 Db 16 AGGCTACTGTGGTTAGTACCTTCTCTCGGCTGGTCAAGAAAGGAAAGGATGTC 75
 |||||
 QY 137 tgccttgggggtgcagctccatcttgcaccatcccgctgacagtcacagtcgcccgggt 196
 |||||
 Db 76 TCGGTTGGGGGTGGGCTCCATCTTACACCATCCACTGACAGTCACACGCGCGGGT 135
 |||||
 QY 197 cagttctctgggttccagcgcctctccagaagtgacgcccgtggctcagctctgaagga 256
 |||||
 Db 136 CAGTTTCTCGGTTCCCGAGCCCTCTGCAGAAAGTACGCGCAGTGCCTCTTTGAAGGA 195
 |||||
 QY 257 cttagacagagcacccttcagaagttgttgaactgttagtcgggcccctgcattggaa 316
 |||||
 Db 196 CTTAGACAGAGGACCTTCAGAAAGTTGTTGAACTGCTAGTCGGGCCCTCCATGGGA 255
 |||||
 QY 317 agactgcagagaagctgtggagcaacttgggtgccagcgccacacctgtcagaagagcgtct 376
 |||||
 Db 256 AGACTGCAGAGAGGCGGTGCAGCACCTTGTGTCTAGCGCCAACTGTACAGAGGAGCGTCT 315
 |||||
 QY 377 ggcgtctctgtgcgggcacacacacccctgtctccagcaggtctcggctgcacctgc 436
 |||||
 Db 316 GGAGCTCTCTGCTGGCGGGCACACACACCTCTCCAGCAGGCTCTCCGGGTGCCCTGTC 375
 |||||
 QY 437 tagtctaaagccagatgccttccaggaagagctccaggaacttgacattctcaggatct 496
 |||||
 Db 376 CAGTCTGAACACAGATGCCCTTCAGAGGACGAACTCCAGAACTTGGCATTCGCCAAGATAT 435
 |||||
 QY 497 aattggagatttggcagattt 517
 |||||
 Db 436 GATTGGAGATCTGGCCAGTTT 456
 |||||

RESULT 5

AA136965/c
 LOCUS UI-R-C2p-nz-d-10-0-UI.s1 UI-R-C2p Rattus norvegicus cDNA clone
 DEFINITION UI-R-C2p-nz-d-10-0-UI 3', mRNA sequence.
 ACCESSION AA136965
 VERSION AA136965.1 GI:3637742
 KEYWORDS EST.
 SOURCE Norway rat.
 ORGANISM Rattus norvegicus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 REFERENCE 1 (bases 1 to 386)
 AUTHORS Bonaldo, M.F., Lennon, G. and Soares, M.B.
 TITLE Normalization and subtraction: two approaches to facilitate gene
 discovery
 JOURNAL Genome Res. 6 (9), 791-806 (1996)
 MEDLINE 97044477
 COMMENT On Jan 19, 1998 this sequence version replaced gi:2150747.
 Contact: Soares, MB
 Program for Rat Gene Discovery and Mapping
 University of Iowa
 451 Eckstein Medical Research Building Iowa City, IA 52242, USA
 Tel: 319 335 8250
 Fax: 319 335 9565
 Email: msoares@blue.weeg.uiowa.edu

The sequence tag present in the cDNA between the NotI site and the
 oligo-dr track served to identify it as a clone from the normalized
 adult Spleen library. cDNA Library Preparation: M. Fatima Bonaldo,
 Ph.D. Clone distribution: clones will be available through Research
 Genetics This clone is also available through the I.M.A.G.E.
 Consortium at LNL (info@image.llnl.gov). IMAGE ID=1787109 The
 following repetitive elements were found in this cDNA sequence:
 33-145, >PBID7#SINE/Alu
 Seq primer: M13 Forward
 POLYA=NO.
 Location/Qualifiers
 1. 386
 /organism="Rattus norvegicus"

FEATURES

source
 1. 386
 /organism="Rattus norvegicus"


```

/strain="Sprague-Dawley"
/db_xref="taxon:10116"
/clone="UI-R-C2p-nz-d-10-0-UI"
/clone_lib="UI-R-C2p"
/dev_stage="adult"
/lab_host="DH10B"
/Note="Vector: pT73D-Pac (Pharmacia) with a modified
polylinker; Site_1: Not I; Site_2: Eco RI; The UI-R-C2p
library is a subtracted library derived from the UI-R-C1
library, which is a subtracted library derived from the
UI-R-C0 library. The UI-R-C0 library consisted of a
mixture of individually tagged normalized libraries
constructed from rat placenta, adult lung, brain, liver,
kidney, heart, spleen, ovary, muscle, 8, 12 and 18-day
embryo. The tag is a string of 3-5 nucleotides present
between the Not I site and the oligo-dT track which allows
identification of the library of origin of a clone within
the mixture. The subtracted library (UI-R-C2p) was
constructed as follows: PCR amplified cDNA inserts from
UI-R-C1 clones from which 3' ESTs had been derived was
used as a driver in a hybridization with the UI-R-C1
library in the form of single-stranded circles. The
remaining single-stranded circles (subtracted library) was
purified by hydroxyapatite column chromatography,
converted to double-stranded circles and electroporated
into DH10B bacteria (Life Technologies) to generate the
UI-R-C2p library. This procedure has been previously
described (Bonaldi, Lennon and Soares, Genome Research 6:
791-806, 1996)"
BASE COUNT      87 a 113 c 71 g 115 t
ORIGIN

Query Match      34.8%; Score 382.8; DB 42; Length 386;
Best Local Similarity 99.5%; Pred. No. 2.5e-97;
Matches 384; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 715 ccaatttcaggagcgcgtacagttagccttggtccttaaggagatggcagaactgg 774
Db 386 CCAAAATTCAGGAGCTCGGTACAGTGTAGCCTTGGTCTTAAGGAGATGGCAGAATGG 327
QY 775 agaagaagtgaacgaactcaggactgactgaacctggtactgtgggtgcgaag 834
Db 326 AGAAGAAAGTGTAGCGCAAACTCGAGGACTGACTGAACCTGGTACTGTGGGTGCTGAAG 267
QY 835 ctggtaccagaacacagcccccactgtgtgatgagcccaactcattgaggtcctgcatg 894
Db 266 CTGGTACCAGAACACAGCCCCCCTGCTGTGATGAGGCCAACTCCATGAGGTCTGCAATG 207
QY 895 tgagaacgtatttaagtgaagacagcgggactttcaggtttttgtttatgagtcac 954
Db 206 TGAGAACGTATTTAAGTGAAGAAAGACAGCGGGACTTTTCAGGTTTTGTTTATGAGTCAAC 147
QY 955 agctggcaggggtggcacatttataatcagcccttggaactctgaggtcggagaatg 1014
Db 146 AGCTGGCAGGGTGGCACATTTATATATCATAGTGAAGCTTGAAGGCTGAGGCTGAGAATG 87
QY 1015 ggaagttaagctgggctggcttcattagtgaggtcaggtctgcatgaaagaggttaa 1074
Db 86 GGAAGTGAAGTGGGCTCGCTTTTCAATAGTGAAGCTCAGGTGCGAATTAAGAGGTTAA 27
QY 1075 gcaactataaaaaaiaaaaaa 1100
Db 26 GCAACTATTAATAAAAAAIAAAAAA 1

RESULT 6
LOCUS      AI048873      482 bp      mRNA      EST      08-JUL-1998
DEFINITION uc76f02.y1 Sugano mouse liver mlia Mus musculus cDNA clone
IMAGE:1431579 5', mRNA sequence.
ACCESSION  AI048873
VERSION     AI048873.1 GI:3297160

```

KEYWORDS SOURCE ORGANISM

EST.
house mouse.
Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE AUTHORS

1 (bases 1 to 482)
Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
Schellenberg,K., Saeptoe,M., Tan,F., Underwood,K., Moore,B.,
Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
Waterston,R.

TITLE JOURNAL COMMENT

The WashU-HMI Mouse EST Project
Unpublished (1996)
On Jan 14, 1998 this sequence version replaced gi:1798596.
Contact: Marra M/Mouse EST Project

WashU-HMI Mouse EST Project Washington University School of Medicine 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108 Tel: 314 286 1800 Fax: 314 286 1810

Email: mouseest@watson.wustl.edu
This clone is available royalty-free through LLNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
MG1:915647

Seq primer: custom primer used High quality sequence stop: 463. Location/Qualifiers

1. 482
/organism="Mus musculus"
/strain="C57BL"
/db_xref="taxon:10090"
/clone="IMAGE:1431579"
/clone_lib="Sugano mouse liver mlia"
/sex="female"
/dev_stage="adult"
/lab_host="DH10B"

FEATURES source

/note="Organ: liver; Vector: pME18-FL3; Site_1: DraIII
(CACTGTGTG); Site_2: DraIII (CACCATGTG); 1st strand cDNA
was primed with an oligo(dT) primer
[ATGGGCGCTTTTGTGTTTGTGTTT]; double-stranded cDNA was
ligated to a DraIII adaptor [GTGTGGCTACTGTTG], digested
and cloned into distinct DraIII sites of the pME18-FL3
vector (5' site CACTGTGTG, 3' site CACCATGTG). XhoI should
be used to isolate the cDNA insert. Size selection was
performed to exclude fragments <1.5kb. Library
constructed by Dr. Sumio Sugano (University of Tokyo
Institute of Medical Science). Custom primers for
sequencing: 5' end primer CTCTGCTCTTAAAGCTGGG and 3' end
primer GACCTGCGCTCGAGCACA."

BASE COUNT ORIGIN

102 a 144 c 136 g 100 t

Query Match Best Local Similarity Matches 400; Conservative 0; Mismatches 39; Indels 0; Gaps 0;

34.2%; Score 376.6; DB 41; Length 482;
91.1%; Pred. No. 1.5e-95;

QY 71 ttgttaatgaactgtgttagacacctctctcgcgactggtcaagaaacggaagaag 130 Db 44 TTCTTTTGTACTGTGGTTAGTACCTTCTCCGGGCTGTATCAAGAAAGGAAAGGAAG 103

131 gatgtctgctttggggtcgagctccactctgacacctcccgctgacagtcacagtgg 190
Db 104 GATGCTGCGTTGGGGCTGCGGCTCCATACCTTACACCATCCCACTGACAGTCACAGCGG 163

QY 191 ccgggtcagtttctctggtttccagccctctccagaaagtcagcggcgtgagctcctt 250 Db 164 CCGGGTCAGTTTCTCTGGGTTCCAGGCCCTCTGCAGAAAGTACAGCGCAGTGGCCAGCTCTT 223

QY 251 gaaggacttagacagcagcacccttcagaagtgttgaaacttgtagtcggggccctgca 310 Db 224 GAAGGACTTAGACAGGAGCAGCTTCAGAAAGTTGTTGAAACTCGTAGTCGGGGCCCTCCA 283

```

QY 311 tgggaagactcagagaagctgtgtgagcaacttgggtccagcgccaaacctgtctcagaaga 370
|||||
Db 284 TGGGAAGACTCGAGAGAGCGCTGACGACCTTGGTGGCAGCGCAACCTGTACAGAGA 343
|||||
QY 371 gcgtggtccgtccctgtggtgggggacacacacacacctgtctcagcagggctctccggtccc 430
|||||
Db 344 GCGTGTGCGCGTCCCTGCTGTGGCGGGGACACACACCTGTCTCCAGCAGGCTCTTCGCGCTGCC 403
|||||
QY 431 cccgtgtctaaagccagatgccttccaggaaagctccaggaagctggcattccctca 490
|||||
Db 404 CCTGCCAGTCTGAACAGATGCTTCCAGGACCACTCCAGGAACTTGGCATTTCCCA 463
|||||
QY 491 ggatctaattggagatttg 509
|||||
Db 464 AGATATGATTGGAGATCTG 482

RESULT 7
AA032667 444 bp mRNA EST 22-AUG-1996
LOCUS mi32b07.r1 Soares mouse embryo NDME13.5 14.5 Mus musculus cDNA
DEFINITION clone IMAGE:465205 5', mRNA sequence.
ACCESSION AA032667
VERSION AA032667.1 GI:1505234
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 444)
AUTHORS Marra, M., Hillier, L., Allen, M., Bowles, M., Dietrich, N., Dubuque, T.,
Geisler, S., Kucaba, T., Lacy, M., Le, M., Martin, J., Morris, M.,
Schellenberg, K., Steptoe, M., Tan, F., Underwood, K., Moore, B.,
Theising, B., Wylie, T., Lennon, G., Soares, B., Wilson, R. and
Waterston, R.
TITLE The WashU-HMI Mouse EST Project
JOURNAL Unpublished (1996)
COMMENT On Nov 29, 1993 this sequence version replaced gi:636534.
Contact: Marra M/Mouse EST Project
WashU-HMI Mouse EST Project
Washington University School of MedicineP
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: mouseest@watson.wustl.edu
This clone is available royalty-free through LLNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
MGI:279021
Seq primer: -28M13 rev2 from Amersham
High quality sequence stop: 434.
FEATURES
source
1. .444
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="IMAGE:465205"
/tissue_type="unknown"
/dev_stage="13.5-14.5dpc total fetus"
/lab_host="DH10B"
/note="vector: pT73D-Pac (Pharmacia) with a modified
polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
was primed with a Not I - oligo(dT) primer [5',
TGTACCAATCTCAAGTGGAGCGCGCGCAATTTTITTTTTTTTTTTT
T 3'], on equal amounts of mRNA from 2 13.5dpc and 2
14.5dpc embryos [total RNA provided by Minoru Ko, Wayne
State Univ., from 2 ]; double-stranded cDNA was ligated to
Eco RI adaptors (Pharmacia), digested with Not I and
cloned into the Not I and Eco RI sites of the modified
pT73 vector. Library went through one round of
normalization, and was constructed by Bento Soares and
M.Fatima Bonaldo."

```

```

BASE COUNT 95 a 133 c 128 g 88 t
ORIGIN
Query Match 33.9%; Score 373; DB 27; Length 444;
Best Local Similarity 91.8%; Pred. No. 1.5e-94;
Matches 394; Conservative 0; Mismatches 35; Indels 0; Gaps 0;
QY 77 atgcaactgtgttaggaccttcttcggactgttcaagaacgggaagaagatgtc 136
|||||
Db 16 AGGTACTGTGTGTAGTACCTTCTCTCGGCTGGTCAAGAAAGGGAAGGATGTC 75
|||||
QY 137 tgccttgggggtcagctccatacttgcaccatcccgctgacacagtcacagtgccgggt 196
|||||
Db 76 TGCCTTGGGGCTCGGGCTCCATCTTACACCATCCAACGTACAGTCCACAGGCGCGGT 135
|||||
QY 197 cagttctcgggtccagcgcctctccagaagtgacggccgtgctcagctcttgaagga 256
|||||
Db 136 CAGTTTCTTGGGTTCGAGCCCTCTGCAGAGTGTGCGAGCGCCAGTGGCCAGCTCTTGAAGGA 195
|||||
QY 257 cttagacagagcaccctcagaaagtgttgaactgttagtcggggccctgcattggaa 316
|||||
Db 196 CTTACACAGGAGCACCTTCAGAAAGTTGTGAAGTCTGTAGTCGGGGCCCTTCCATGGGA 255
|||||
QY 317 agactgcagagaagctgtgagcaacttgggtgccagcgccaaacctgtcagaagagctct 376
|||||
Db 256 AGACTGCAGAGAGCGCGCTGCAGCACCTTGTGTCAGCGCCCAACCTGTGCAGAGAGCGCTCT 315
|||||
QY 377 ggcgtctcgtcgtgggggacacacacacctgtctccacaggtctccggtgcacctgc 436
|||||
Db 316 GGCGTCTCTGCTGGCGGGCACACACACCTTGTCTCCAGCAGGCTCTCGGGTGGCCCTGTC 375
|||||
QY 437 tagtctaaagccagatgcctccaggaagagctccaggaacttgcattctcaggatct 496
|||||
Db 376 CAGTCTGAACACAGATGCTTCCAGGACGAACTCCAGAACTTGGCAATTTCCCAAGATAT 435
|||||
QY 497 aattggaga 505
|||||
Db 436 GATTGGAGA 444

```

```

RESULT 8
AA154070 454 bp mRNA EST 30-SEP-1998
LOCUS ud76a06.r1 Soares_mammary_gland_NMLMG Mus musculus cDNA clone
DEFINITION IMAGE:1476754 5', mRNA sequence.
ACCESSION AA154070
VERSION AA154070.1 GI:3682539
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 454)
AUTHORS Marra, M., Hillier, L., Allen, M., Bowles, M., Dietrich, N., Dubuque, T.,
Geisler, S., Kucaba, T., Lacy, M., Le, M., Martin, J., Morris, M.,
Schellenberg, K., Steptoe, M., Tan, F., Underwood, K., Moore, B.,
Theising, B., Wylie, T., Lennon, G., Soares, B., Wilson, R. and
Waterston, R.
TITLE The WashU-HMI Mouse EST Project
JOURNAL Unpublished (1996)
COMMENT On Jan 14, 1998 this sequence version replaced gi:1797639.
Contact: Marra M/Mouse EST Project
WashU-HMI Mouse EST Project
Washington University School of MedicineP
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: mouseest@watson.wustl.edu
This clone is available royalty-free through LLNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
MGI:925110
Seq primer: -28m13 rev2 ET from Amersham

```

FEATURES		High quality sequence stop: 452.		TITLE		Normalization and subtraction: two approaches to facilitate gene discovery	
SOURCE		Location/Qualifiers		JOURNAL		Genome Res. 6 (9), 791-806 (1996)	
		1. .454		MEDLINE		97044477	
		/organism="Mus musculus"		COMMENT		On May 18, 1998 this sequence version replaced gi:3137829.	
		/db_xref="taxon:10090"				Contact: Soares, MB	
		/clone="IMAGE:1476754"				Program for Rat Gene Discovery and Mapping	
		/sex="female (lactating)"				University of Iowa	
		/tissue_type="mammary gland"				451 Eckstein Medical Research Building Iowa City, IA 52242, USA	
		/lab_host="DH10B"				Tel: 319 335 8250	
		/note="Vector: pT73D-Pac (Pharmacia) with a modified polylinker; 1st strand cDNA was prepared from mammary gland tissue from a lactating female, and was then primed with a Not I - oligo(dT) primer. Double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT73 vector. Library is normalized. Library was constructed by Bento Soares and M. Fatima Bonaldo."				Fax: 319 335 9565	
BASE COUNT		96 a 131 c 134 g 93 t				Email: msoares@blue.weeg.uiowa.edu	
ORIGIN						The sequence contained an oligo-dT track that was present in the oligonucleotide that was used to prime the synthesis of first strand cDNA and therefore this may represent a bonafide poly A tail. The sequence tag present in the cDNA between the NotI site and the oligo-dT track served to identify it as a clone from the normalized atrium at 15 dpc library cDNA Library Preparation: M.B. Soares lab clone distribution: clones will be available through Research Genetics (www.resgen.com) The following repetitive elements were found in this cDNA sequence: 34-146, >PB1D7#SINE/Alu	
		Seq primer: M13 Forward					
		POLYA=Yes					
				FEATURES		Location/Qualifiers	
				Source		1. .345	
						/organism="Rattus norvegicus"	
						/strain="Sprague-Dawley"	
						/db_xref="taxon:10116"	
						/clone="UI-R-BJ0p-afi-e-01-0-UI"	
						/clone_lib="UI-R-BJ0p"	
						/dev_stage="adult"	
						/lab_host="DH10B (Life Technologies)"	
						/note="Vector: pT73D-Pac (Pharmacia) with a modified polylinker; Site 1: Not I; Site 2: Eco RI; The UI-R-BJ0p library is a subtracted library derived from the UI-R-AAL, UI-R-ABL, UI-R-ACL, UI-R-ADL, UI-R-AEL, UI-R-AFI, and UI-R-AGL libraries. These libraries represent tissues from rat atrium at 16.5 dpc, ventricle at 16.5 dpc, AV canal at 16.5 dpc, atrium at 15 dpc, ventricle at 15 dpc, AV canal at 15 dpc, and ventricle at 13 dpc. The tag is a string of 5-6 nucleotides present between the Not I site and the oligo-dT track. The library was constructed as described by Bonaldo, Lennon and Soares, Genome Research 6: 791-806, 1996.	
						TAG_LIB=UI-R-BJ0p	
						TAG_TISSUE=atrium at 15 dpc	
BASE COUNT		77 a 101 c 63 g 104 t					
ORIGIN						TAG_SEQ=GAAGC"	
				Query Match		30.4%; Score 334.4; DB 81; Length 345;	
				Best Local Similarity		98.3%; Pred. No. 1e-83;	
				Matches 338; Conservative		0; Mismatches 6; Indels 0; Gaps 0;	
Qy		757 agagatggcagaactggagaagaagtgtgagcgcaaatcaggactgactgaacctg 816					
Db		345 AGGAGATGGCAGAACTGGAGAAGTGTGAGCGCAAACTCCAGGACTGACTGAACCTG 286					
Qy		817 gtactgtgggtgtgtaagctggtaccagaacacagccccccactggtgatgagcccaact 876					
Db		285 GTACTGTGGGTGTGTAAGCTGGTACCAGAACACAGCCCCCTGTTGATGAGGCCAACT 226					
Qy		877 ccattgaggtccctgcattgtgagaactgtattttaagtaaaagacagcgagacttcaggt 936					
Db		225 CCATTGAGGTCCCTGCATGTGAGAACGTATTTTAAAGTAAAGACAGCGGACATTTCAGGT 166					
Qy		937 ttgtttttatgagtcacacagctgggcagggtgtggcacagttataatctcgcccttgaa 996					
Db		165 TTGTCTTTATGAGTCAACAGCTGGCGAGGGTGGCACAGTTTATAATCTCAGCCCTTGAA 106					
Qy		997 gtctgaggtcggagaatggaagaagtgaagctggccttcggttttcattcatagtaggctcagtg 1056					


```
COMMENT      On Apr 18, 1995 this sequence version replaced gi:775302.
Contact: Hirofumi Doi
Doi Biosymmetry Project, ERATO
Japan Science and Technology Corporation (JST)
WBG Marive East 12F, 2-6 Nakase, Mihama-ku, Chiba 261-71, Japan
Email: hdo@boa.jst.go.jp.

FEATURES
Source      Location/Qualifiers
1..569
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="J0040D09"
/clone_lib="Mouse 3.5-dpc blastocyst cDNA"
/tissue_type="blastocyst"
/dev_stage="3.5-dpc"

BASE COUNT  133 a 169 c 131 g 134 t 2 others
ORIGIN

Query Match      27.0%; Score 296.6; DB 36; Length 569;
Best Local Similarity 80.6%; Pred. No. 5.6e-73;
Matches 510; Conservative 0; Mismatches 46; Indels 77; Gaps 11;

QY 460 aggaagagctccaggaaacttggtcattctc--aggatctaattgagatttgccagttt 517
      |||||
DB 569 AGGACGAACCTCCAGGAACCTGGCAATTCCTCCCAAGATTATGATGGAGATCTGGCCAGTTT 510
      |||||

QY 518 ggcatttgaggagtaacagccc-tctctcgaactctgtagcccaacagcaggat-cctcg 575
      |||||
DB 509 GGCATTTGGGANTCAACGCCCTTTTCTGACTCTGTAGCCAGCAGCAGGATCCCTCG 450
      |||||

QY 576 ctgcctcaagtgtcttactcctggcgggtggagctggcagctggccatctcaaccagcgtcag 635
      |||||
DB 449 CTGCCTCGGTGTCGAAATTCGGGTGGCGGGTGGATGGCCATCTCAACCAAGTCTCAG 390
      |||||

QY 636 tccgcctcctcgaaccagagtggttctcatg-cagctgaagctcacagatgagatctgca 694
      |||||
DB 389 TCCCGCTCCCTCAACACCGAGTGTCTCATGTCAGCTGAAGCTCACAGATGATGGGCACA 330
      |||||

QY 695 ccgctctcaggtgcccataagcacaattccagagctgcggtacagtgtagccttggtct 754
      |||||
DB 329 CGCCTTCAGGTGCCCATAGCCAAATTCAGAGAGTGTCTACAGAGTGTCTACAGCGTGGTGCCT 270
      |||||

QY 755 taagagagtgccagaaactggagaagaagtgtgagcgcgaactgagagctgactgaaccc 814
      |||||
DB 269 TAAGGAGATGGCAGAGCTGGAGAGAAAGTGTGAGCGCAAACTGCAG---GACTGAACCC 214
      |||||

QY 815 tggtagctgt-gggtgctgaagctggtaccagaacacagcccccaactggtgtagagccca 873
      |||||
DB 213 TGGTACTGTGGGTGCTGAAGCTGGTACCAGAACCCAG--CCGCACTGGTGTATGA----- 160
      |||||

QY 874 actccattgaggtcctgcgcatgtgagaacgtattttaagtgaagaaagacagc-gggactttc 932
      |||||
DB 160 -----GGATCTAAGACATATATTTAAGTGAAGAAAGACAGCTGGGATTTT 117
      |||||

QY 933 aggtttgttttatgagtcacagctgggaggggtgggcagagtttataatctcagccctt 992
      |||||
DB 116 GGGTTTGTTTTACAAATAAACAGCTGGG-----87
      |||||

QY 993 ggaagcttgagcctggagaatgggaagtgtgaagctggaagctggcctggc----- 1046
      |||||
DB 87 -----CTGAGGCTGGAGATGTGAATGTGAAGCGCGCGCTGGCTGGTGTACCTAGTGTG 34
      |||||

QY 1047 aggtcagtgctcgaattaaagaggtaaagcaac 1079
      |||||
DB 33 AGGCTCAGTTCGAATTAAGAGAGTAAAGCAAC 1
      |||||

RESULT 14
AA450933
LOCUS
DEFINITION      374 bp, mRNA
vfi85all.r1 Soares_mammary_gland_NBMWG Mus musculus cDNA clone
IMAGE:850556 5', mRNA sequence.

ACCESSION      AA450933
VERSION      AA450933.1
KEYWORDS      GI:2164603
SOURCE      house mouse.
ORGANISM      Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 374)
Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
Geisler,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
Waterston,R.
The WashU-HMI Mouse EST Project
Unpublished (1996)
On Sep 12, 1996 this sequence version replaced gi:1394992.
Contact: Marra M/Mouse EST Project
WashU-HMI Mouse EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: mouseest@watson.wustl.edu
This clone is available royalty-free through LLNL: contact the
IMAGE Consortium (image.llnl.gov) for further information.
MGI:502708
Seq primer: -28ml3 rev2 ET from Amersham
High quality sequence stop: 348.
Location/Qualifiers
1..374
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="IMAGE:850556"
/clone_lib="Soares_mammary_gland_NBMWG"
/tissue_type="mammary gland"
/dev_stage="4 weeks"
/lab_host="DH10B"
/note="Organ: mammary gland; Vector: pT73D-Pac
(Pharmacia) with a modified polylinker; Site 1: Not I;
Site 2: Eco RI; 1st strand cDNA was primed with a Not I -
oligo(dT) primer [5',
TGTACCAATGTAAGTGGAGCGCGCGCAATGTTTGTGTGTGTGTGTGTGTGTGT
T 3']; double-stranded cDNA was ligated to Eco RI
adaptors (Pharmacia), digested with Not I and cloned into
the Not I and Eco RI sites of the modified pT73 vector.
RNA provided by Dr. Minoru Ko, Wayne State Univ. Library.
Constructed and normalized by Bento Soares and M.Fatima
Bonaldo."
BASE COUNT      77 a 113 c 112 g 72 t
ORIGIN

Query Match      26.9%; Score 296.4; DB 33; Length 374;
Best Local Similarity 92.3%; Pred. No. 5.5e-73;
Matches 312; Conservative 0; Mismatches 26; Indels 0; Gaps 0;

QY 77 atgcaactgtggttaggccttctcttcgagctgttcacgaacacgggaagagatgctc 136
      |||||
DB 37 AGGCTACTGTGGTAGTACCTTCTCTCGGCTGTCTCAAGAAAGGAAAGAGAGATGTC 96
      |||||

QY 137 tgccttggggggtcagctccatcttcgacctcccgctgacagtcacagtcagtcgctc 196
      |||||
DB 97 TCGGTGGGGGCTCGGGCTCCATCTTACACCATCCAACTGACAGTACAGCGCCCGGT 156
      |||||

QY 197 cagtttctcctgggttcccgagccctctccagagtgacggccgtggtcagctcttgaagga 256
      |||||
DB 157 CAGTTTCCTGGGTTCGCCGCCCTCTGCAGAGTGCAGCGCAGTCCAGCTCTTGAAGGA 216
      |||||

QY 257 cttagacaggagcaccttcagaaagtgttgaaactgttagtcggggccctgcatgggaa 316
      |||||
DB 217 CTTAGACAGGAGCACCCTCAGAAAGTTGTTGAARACTCGTAGTCGGGGCCCTCCATGGGAA 276
      |||||
```

Search completed: May 14, 2000, 23:17:58
Job time: 16358 sec

Query Match	26.1%	Score 286.6;	DB 36;	Length 574;
Best Local Similarity	80.3%	Pred. No. 3.7e-70;		
Matches 500;	Conservative 0;	Mismatches 46;	Indels 77;	Gaps
Qy 470	ccaggaacttggattctcctcagatctaattggagatttggccagattgg--catttggg	527		
Db 559	CCAGGAACITGGCATTCCCCANATGATGTGGAGATCTGGCCCAAGTTGGCCCAITGGG	500		
Qy 528	agtcacagccctcttcgcactctgtag--cccaacagcaggat--cctgcgtgcctcacg	585		
Db 499	AGTCACAGCCNCTTCTTCGACTCTGTAGCCCCAGCAGCAGGGATCCCTCGCTGCCTCGTG	440		
Qy 586	tgtcttacttcctcgggtggcgggtggacgtggccatctctcaaccagcgcctcagtcctccgcctccc	645		
Db 439	TGTCGAATTTCCGGTGGCGGGTGGATGTGGCCCATCTCAACCAAGTGTCTAGTCCCGCTCCC	380		
Qy 646	tgcacccagatgttctctatg--cagctgaagctcacagatggtatctgtcacacgccttcag	704		
Db 379	TGCAACCCAGTGTCTCATGCCAGCTGAAGCTCACAGATGGATTCGGCACACCGCTTCGAG	320		
Qy 705	gtgccctatagccaaatttcaggagctgcggtacagtgtagccttggcttccttaagagatg	764		
Db 319	GTGCCCATAGCCAAATTCACAGGAGCTGCTCTACAGCGTGGCTTGGTCTCTTAAGGAGATG	260		
Qy 765	gcagaactggagaagaagtgtgagcgcacaaactcgaggactgactgaacctgggtactgt--	824		

" "

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model
Run on: May 14, 2000, 23:17:58 ; Search time 871.67 Seconds
(without alignments)
6300.684 Million cell updates/sec

Title: US-09-223-796-3
Perfect score: 1355
Sequence: 1 999caggcagtgagggtgga.....gtgttcaggcaggcccg 1355

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 4857316 seqs, 2026611650 residues
Total number of hits satisfying chosen parameters: 9714632
Minimum DB seq length: 0
Maximum DB seq length: 1000000
Post-processing: Minimum Match 0%
Listing first 45 summaries

Database :		EST:*
1:	em_est1:*	45: gb_est26:*
2:	em_est2:*	46: gb_est27:*
3:	em_est3:*	47: gb_est28:*
4:	em_est4:*	48: gb_est29:*
5:	em_est5:*	49: gb_est30:*
6:	em_est6:*	50: gb_est31:*
7:	em_est7:*	51: gb_est32:*
8:	em_est8:*	52: em_est20:*
9:	em_est9:*	53: em_est21:*
10:	em_est10:*	54: em_est22:*
11:	em_est11:*	55: em_est23:*
12:	em_est12:*	56: em_est24:*
13:	em_est13:*	57: em_est25:*
14:	em_est14:*	58: em_est26:*
15:	em_est15:*	59: gb_est33:*
16:	em_est16:*	60: gb_est34:*
17:	em_est17:*	61: gb_est35:*
18:	em_est18:*	62: gb_est36:*
19:	em_est19:*	63: gb_est37:*
20:	gb_est1:*	64: gb_est38:*
21:	gb_est2:*	65: em_est27:*
22:	gb_est3:*	66: em_est28:*
23:	gb_est4:*	67: em_est29:*
24:	gb_est5:*	68: em_est30:*
25:	gb_est6:*	69: gb_est39:*
26:	gb_est7:*	70: gb_est40:*
27:	gb_est8:*	71: gb_est41:*
28:	gb_est9:*	72: gb_est42:*
29:	gb_est10:*	73: gb_est43:*
30:	gb_est11:*	74: gb_est44:*
31:	gb_est12:*	75: em_est31:*
32:	gb_est13:*	76: em_est32:*
33:	gb_est14:*	77: em_est33:*
34:	gb_est15:*	78: em_est34:*
35:	gb_est16:*	79: gb_est45:*
36:	gb_est17:*	80: gb_est46:*
37:	gb_est18:*	81: gb_est47:*
38:	gb_est19:*	82: gb_gss1:*
39:	gb_est20:*	83: gb_gss2:*
40:	gb_est21:*	84: gb_gss3:*
41:	gb_est22:*	85: gb_gss4:*
42:	gb_est23:*	86: em_gss1:*
43:	gb_est24:*	87: em_gss2:*
44:	gb_est25:*	88: em_gss3:*
		89: em_gss4:*
		90: gb_gss5:*
		91: gb_gss6:*
		92: gb_gss7:*
		93: gb_gss8:*
		94: gb_gss9:*
		95: em_gss5:*
		96: em_gss6:*
		97: em_gss7:*
		98: em_gss8:*
		99: em_gss9:*
		100: em_gss10:*
		101: em_gss11:*
		102: gb_gss10:*
		103: gb_gss11:*
		104: em_gss12:*
		105: gb_gss12:*
		106: gb_gss13:*
		107: gb_gss14:*
		108: gb_gss15:*
		109: gb_gss16:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result	%	Query
--------	---	-------


```
QY 1167 cagct 1171
Db 5 CAGCT 1

RESULT 2
LOCUS AW005581/c
DEFINITION w287h11.x1 NCI_CGAP_Brn25 Homo sapiens cDNA clone IMAGE:2565861 3',
mRNA sequence.
ACCESSION AW005581
VERSION AW005581.1 GI:5854359
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 520)
AUTHORS NCI/NINDS-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute / National Institute of Neurological
Disorders and Stroke Brain Tumor Genome Anatomy Project
(CGAP/BRGAP), Tumor Gene Index
JOURNAL Unpublished (1998)
COMMENT On Jun 5, 1998 this sequence version replaced gi:3187080.
Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert.Strausberg@nih.gov
Tissue Procurement: David N. Louis, M.D., Myrna R. Rosenfeld M.D.,
Ph.D.
CDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima
Bonaldo, Ph.D.
CDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www.bio.llnl.gov/bbrp/image/image.html

Seq primer: -40UP from Gibco
High quality sequence stop: 452.
FEATURES
source
Location/Qualifiers
1..520
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:2565861"
/clone_lib="NCI_CGAP_Brn25"
/tissue_type="anaplastic oligodendroglioma"
/lab_host="DH10B"
/notes="Organ: brain; Vector: pT7T3D-Pac (Pharmacia) with a
modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st
strand cDNA was primed with a Not I - oligo(dT) primer [5'
TGTTACCATCTCAAGTGGGAGCGCGCATAGTTTCTTTTCTTTTCTTTTCTTTT
T 3']; double-stranded cDNA was ligated to Eco RI
adaptors (Pharmacia), digested with Not I and cloned into
the Not I and Eco RI sites of the modified pT7T3 vector.
Library is normalized, and was constructed by Bento
Soares and M.Fatima Bonaldo."
BASE COUNT 124 a 150 c 137 g 109 t
ORIGIN

Query Match 37.7%; Score 510.8; DB 63; Length 520;
Best Local Similarity 99.6%; Pred. No. 1.1e-117;
Matches 512; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 662 tcggggacttgccagcgtggtatttggagccagcggccctcttgattctgtggccc 721
Db 520 TCGGGGACTTGGCCAGCGTGGTATTGGGAGCCAGCGCCCTCTTGATTCTGTGGCCC 461

QY 722 agacagaggggctgctgcccatcttctgacttcggtggcggtggatgaacaa 781
Db 460 AGCAGCAGGGGGCTGGCTGCCGATCTTGTGACTTTCGGTGGCGGGTGGATGACAA 401
```

```
QY 782 tctccaccagtcgctgctcctccctcagccagcgcctcctgatgcagctgaagcttt 841
Db 400 TCTCCACCAAGTGCCTGCTGCTCCTCCCTCAGCCGAGCGCTCTGTATGTCAGCTGAAGCTTT 341

QY 842 cagatgggtcagcataccgccttgaggtcccccacacagcagccaagttccaggagctgcggtaca 901
Db 340 CAGATGGGTGAGCATACCGCTTTGAGGTCCCCACACAGCCAAGTTCCAGGAGCTGCGGTACA 281

QY 902 ggtggcccttgctcctaaaggagatggcagatcctggagaagaggtgtgagcagcagctgc 961
Db 280 GGTGGCCCTTGCTCTCTAAAGGAGATGGCAGATCTGGAGAAGAGGTGTGAGCGCAGACTGC 221

QY 962 aggaactgacccctcacttgaccagctccctcagatcccggtctggacagcagccactgagat 1021
Db 220 AGGACTGACCCCTCACTTGACCAGTCCCATTCAGATCCCGCTTGGACAGCCACTGAGAT 161

QY 1022 ggtgcaaaagtgcagctgactcttccacagcagccctggcctccctccatcagcagcgcctc 1081
Db 160 GGTGCAAAAGTGCAGCTGACTCTTCCACAGCAGACGCCCTGCCCTTCCCATGAGCAGGCTC 101

QY 1082 ttcaagtgaagttgaacgttaattatgtatgtttctgttttaattgaaaaagagagctatg 1141
Db 100 TTCAGTGAAGTGTGAACTGAATATGTAGTTTCTGTGTTAATTGAAAAGAGAGCTATG 41

QY 1142 cctttttcttttttggaaagtaagcagctaaaa 1175
Db 40 CCTTTTCTTTTGGAAAGTAAGCAGCTAAAA 7

RESULT 3
LOCUS AW0055548/c
DEFINITION CIT-HSP-2338P5.TR CIT-HSP Homo sapiens genomic clone 2338P5,
genomic survey sequence.
ACCESSION AW0055548
VERSION AW0055548.1 GI:3352154
KEYWORDS GSS.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 540)
AUTHORS Adams,M.D., Rounsley,S.D., Zhao,S., Field,C.E., Bass,S., Linher,K.,
Golden,K., Berry,K., Granger,D., Suh,E., Wible,C., Shizuya,H.,
Simon,M. and Venter,J.C.
TITLE Use of a random BAC End Sequence Database for Sequence-Ready Map
Building (1998)
JOURNAL Unpublished (1998)
COMMENT Other_GSSs: CIT-HSP-2338P5.TF
Contact: Mark Adams
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: mdadams@tigr.org
Clones are available from Research Genetics (info@resgen.com). BAC
end search page:
http://www.tigr.org/tldb/humgen/bac_end_search/bac_end_search.html.
Seq primer: M13 Reverse
Class: BAC ends.
FEATURES
source
Location/Qualifiers
1..540
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="2338P5"
/clone_lib="CIT-HSP"
/sex="Male"
/cell_type="Sperm"
/notes="Vector: pBelobAC11; Site_1: HindIII; Site_2:
HindIII"
BASE COUNT 107 a 170 c 166 g 97 t
ORIGIN
```


[illegible]

Clone distribution: NCI-CCAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: www-bio.llnl.gov/pbrp/image/image.html

Insert Length: 1066 Std Error: 0.00
Seq primer: -40UP from Gibco
High quality sequence stop: 404.

	Query Match	30.5%	Score 412.8	DB 43	Length 431	
	Best Local Similarity	99.5%	Pred. No. 3.7e-93			
	Matches 414	Conservative 0	Mismatches 2	Indels 0	Gaps 0	
Qy	760	cggctgcgggtggatgtagcaatctccaccagtcgacctggctgcgtccctgcagccgagc	819			
Db	431	CGGTGCGGGTGGATGTAGCAATCTCCACGAGTCCCTGGCTGCTCCCTGCAGCCGAGC	372			
Qy	820	gtcctgatgcagctgaagctttcagatgggttcagatcacccgttcgtgaggtccccaagcc	879			
Db	371	GTCTGATGCAGCTGAAGCTTTTCAGATGGGTGAGCATACCGCTTTGAGTCCCCACAGCC	312			
Qy	880	aagttccaggagctgcggtacagctggccctggcttcctaaaggagatggcagatcctggag	939			
Db	311	AAGTTCAGAGAGCTGCGGTACAGCTGGCCCTGTGTCTTAAGAGAGATGGCAGATCTGGAG	252			
Qy	940	aagaggtgtgagcgagactgcaggaactgacccctcacttgaccagtcctccattcagatcc	999			
Db	251	AAGAGGTGTGAGCGCAGACTGCAGGACTGACCCCTCACTTGACCAGTCCCATTCAGATCC	192			
Qy	1000	ggcttgacaggcaactggagatgggtgcacaaatgcagctgactcttcccacgacagccct	1059			
Db	191	GGCTTGGACAGGCAACCTGAGATGGTGCCAAAGTGAGCTGACTCTTCCACGAGAGCCCT	132			
Qy	1060	ggccttcccactcagcgagggctcttcagtgagtggttgaaacgtaattatgtagtttctgt	1119			
Db	131	GCCCTTCCCATGAGGCAGGCTCTTCAGTGAGTGGTTGAACGTAAATATGATGATTTCTGT	72			
Qy	1120	ttaattgaaaaagagagctatgcctctttttcttttttggagtgaaagcagctaaaa	1175			
Db	71	TTAATTGAAAAGAGAGCTATGCTTTTTTTCTTTTTTTTGAAGTAAGCAGCTAAAA	16			

RESULT	12
LOCUS	A1076231/c
DEFINITION	A1076231 413 bp mRNA EST 24-SEP-1998 OY65C09.x1 NCI_CGAP_CLL1 Homo sapiens cDNA clone IMAGE:1670704.3', mRNA sequence.
ACCESSION	A1076231
VERSION	A1076231.1 GI:3405409
KEYWORDS	EST.
SOURCE	human.
ORGANISM	Homo sapiens
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo. 1 (bases 1 to 413)

RESULT	13
AW261921/c	
LOCUS	xp87f11.x1 NCI_CGAP_Brn53 Homo sapiens cDNA clone IMAGE:2747373 3'
DEFINITION	EST mRNA 411 bp NCBI_CGAP_Brn53 Homo sapiens cDNA clone IMAGE:2747373 3',
ACCESSION	AW261921
VERSION	AW261921.1 GI:6638737
KEYWORDS	EST.
SOURCE	human.
ORGANISM	Homo sapiens
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
AUTHORS	Eutheria; Primates; Catarrhini; Hominidae; Homo.
TITLE	1. (bases 1 to 411)
JOURNAL	NCI/NINDS-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
COMMENT	National Cancer Institute / National Institute of Neurological Disorders and Stroke, Brain Tumor Genome Anatomy Project (CGAP/BTGP), Tumor Gene Index Unpublished (1998) On Jun 5, 1998 this sequence version replaced gi:3189186. Contact: Robert Strausberg, Ph.D. Tel.: (301) 496-1550 Email: Robert_Strausberg@nih.gov Tissue Procurement: Chris Moskaluk, M.D., Ph.D., Michael R. Emmert-Buck, M.D., Ph.D. CDNA Library Preparation: Life Technologies, Inc. CDNA Library Arrayed by: Christa Prange, The I.M.A.G.E. Consortium DNA Sequencing by: Washington University Genome Sequencing Center Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at: www-bio.lnl.gov/bbrp/image/image.html
FEATURES	Possible reversed clone: polyT not found Seq primer: -40UP from Gibco High quality sequence stop: 396. Location/Qualifiers 1..411 /organism="Homo sapiens" /db_xref="taxon:9606" /clone="IMAGE:2747373" /clone_lib="NCI_CGAP_Brn53" /tissue_type="three pooled meningiomas" /lab_host="DH10B" /note="Organ: brain; Vector: pCMV-SPORT6; Site_1: SalI; Site_2: NotI; Cloned unidirectionally. Primer: oligo dr. Library constructed by Life Technologies."
BASE COUNT	99 a 111 c 107 g 94 t
ORIGIN	
Query Match	29.9%; Score 405.8; DB 79; Length 411;
Best Local Similarity	99.5%; Pred. No. 2.1e-91;
Matches	407; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
OY	767 gggtgtagtgagcaatctccaccaggtcgcctggcctcgctcgcctccctgcacgcagagctcga 826
Db	 411 GGGTGGATGTAGCAANTCTCACAGTGCCCTGGCTGCCTCCCTTCGACGCCGAGCGCTCTGA 352
OY	827 tcagctgaagcttccagatgggtcagcatgccgctttgaggtccccacagccaagtccc 886
Db	 351 TCAGCTGAAGCTTTTCAGATGGGTGGCTAGCATACCGCTTTGAGGTCCCCCAGCAAGTTCC 292
OY	887 agpagctgcggtacacgctggccctgttgtctaagaagagatggcagatctggagaagagt 946
Db	 291 AGGAGTCGCGGTACACGCTGGCCCTCGTCTTAAGGAGATGGCAGATCTGGAGAAGAGT 232
OY	947 gtgagcgcagactgcaggagactgacctcacttgaccagctccattccagatccgctgtg 1006
Db	 231 GTGAGCGCAGACTGCAGGACTGACCCCTCACITGACCAGTCCCATTCAGATCGGCTTGG 172
OY	1007 acagcacctgagatggtgtccaaagtgcagctgactcttccccacagacgcttgccttc 1066
Db	 171 ACAGGCACCTGAGTGGTGTCCAAAGTGCAGCTGACTTCTCCACGACACGCTGCGCTTC 112

```

QY 947 gtgagcagactgagactgacctcacttgaccctcactgaccagctccattcagatccggttg 1006
Db 238 GTGAGCCGACACTGCAGAGTGCACCTCTACTTGACCACTGCCATTGAGATCCGGCTGG 179
QY 1007 acaggccacctgagatggtgccaagtgcagctgacttctccacgacagccctggccttc 1066
Db 178 ACAGGCACCTGAGATGTCGCAAGTGCAGTCACTCTCCACGACAGCCCTGCCCTTC 119
QY 1067 ccatacggcaggtcttcaatgagtggtttgaacgtaattatgtagttctgtttaattg 1126
Db 118 CCATAGGCGAGGCTCTTCAGTGAAGTGTTCAGCGTAATTATGATGTTTCTGTTTAATTG 59
QY 1127 aaaaagagactgaccttttttttttttttttttttttttttttttttttttttttttt 1175
Db 58 AAAAAGAGACTAGCTCTTTTCTTTTGGAGTAAGACGACTAAAA 10

RESULT 15
AA857959/c
LOCUS
DEFINITION
  AA857959 410 bp mRNA EST 25-MAR-1998
  oe33q04.s1 NCI_CGAP_Pr25 Homo sapiens cDNA clone IMAGE:1410390 3',
  mRNA sequence.
ACCESSION
  AA857959
VERSION
  AA857959.1 GI:2946261
KEYWORDS
  EST.
SOURCE
  human.
ORGANISM
  Homo sapiens
  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
  Eutheria; Primates; Catarrhini; Hominidae; Homo.
  1 (bases 1 to 410)
  NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
  Tumor Gene Index
  Unpublished (1997)
  On Jan 19, 1998 this sequence version replaced gi:2284841.
  Contact: Robert Strausberg, Ph.D.
  Tel: (301) 496-1550
  Email: Robert.Strausberg@nih.gov
  Tissue Procurement: Suzanne L. Topalian, M.D., Robert K. Bright,
  Ph.D.
  CDNA Library Preparation: Stratagene, Inc.
  CDNA Library Arrayed by: Greg Lennon, Ph.D.
  DNA Sequencing by: Washington University Genome Sequencing Center
  Clone Distribution: NCI-CGAP clone distribution information can be
  found through the I.M.A.G.E. Consortium/LLNL at:
  www-bio.llnl.gov/dbp/image/image.html

Insert Length: 1238 Std Error: 0.00
Seq primer: -40ml3 fwd. Et from Amersham
High quality sequence stop: 401.
Location/Qualifiers
  1. .410
  /organism="Homo sapiens"
  /db_xref="taxon:9606"
  /clone="IMAGE:1410390"
  /clone_lib="NCI_CGAP_Pr25"
  /tissue_type="epithelium (cell line)"
  /lab_host="SOLR (kanamycin resistant)"
  /note="Organ: prostate; Vector: Bluescript SK-; Site_1:
  EcoRI; Site_2: XhoI; Cloned unidirectionally. Primer:
  Oligo df. Normal prostate epithelial cell line (HPV
  immortalized). 5' adaptor sequence: 5' GAATTCGCGACGAG 3'
  3' adaptor sequence: 5' CTCGAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 3'
  Average insert size: 1.1 kb."
  BASE COUNT 97 a 105 c 107 g 101 t
  ORIGIN

```

Query Match 29.3%; Score 396.8; DB 39; Length 410;
 Best Local Similarity 99.5%; Pred. No. 3.7e-89;
 Matches 398; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```

QY 776 taqcaatctccaccagtgccctggctcgctccctgcagcgagcgctcctgatcgagctga 835
Db 410 TAGCAATCTCCACAGTGGCCCTGGCTCGCTCCCTGCAGCCGAGCGTCTGATCAGCTGA 351
QY 836 agctttcagatgggtcagcataccgcttttgaggttccccacagccaagtccaggaagctgc 895
Db 350 AGCTTTTCAGATGGGTCAAGTACCGGTTTGAGGTCCCCACAGCAAGTTCCAGGAGCTGC 291
QY 896 ggtacagcggtggccctggctcctaaaggagatggcagatctgggagaagaggtgtgagcgca 955
Db 290 GGTACAGCGTGGCCCTGGTCTTAAGAGAGATGCGAGATCTGGAGAAAGAGGTGTGAGCGCA 231
QY 956 gactgcagactgagccctcacttgcacagctccattccagatccggcttggacagcgacc 1015
Db 230 GACTGCAGGACTGACCCCTCCTACCTTGACCAGTCCCATTCAGATCCGCTTGGACAGGCACC 171
QY 1016 tgagatgggtgccaaagtgcagctgactcttcccacagacagccctggcctcccatcaggc 1075
Db 170 TGAGATGGTGGCAAAAGTGCAGCTGACTCTTCCACACAGCCCTGCCCTTCCCATGAGGC 111
QY 1076 aggtcttcagtgagtggtttgaacgtaattatgtagtttctgttttaattgaaaaagaga 1135
Db 110 AGGCTCTTCAGTGAGTGTGTAACGTAATTATGTAAGTCTTCTGTTTAAATGAAAAAGAGA 51
QY 1136 gctatgaccttttttttttttttttttttttttttttttttttttttttttttttttt 1175
Db 50 GCTATGCCCTTTTCTTTTGGAGTAAGCAGCTAAAA 11

```

Search completed: May 14, 2000, 23:18:06
 Job time: 16366 sec

" "

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 14, 2000, 22:50:12 ; Search time 76.12 seconds
(without alignments)
1834.863 Million cell updates/sec

Title: US-09-223-796-1

Perfect score: 1100
Sequence: 1 gcacgagccacagccagcta.....attaaaaaaaaaaaaa 1100

Scoring table: IDENTITY_NUC
Gapop 10_0 , Gapext 1.0

Searched: 226296 segs, 63486255 residues

Total number of hits satisfying chosen parameters: 452592

Minimum DB seq length: 0
Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database : Issued Patents_NA.*
1: /cgn2_6/ptodata/2/ina/5A_COMB.seq.*
2: /cgn2_6/ptodata/2/ina/5B_COMB.seq.*
3: /cgn2_6/ptodata/2/ina/5C_COMB.seq.*
4: /cgn2_6/ptodata/2/ina/5D_COMB.seq.*
5: /cgn2_6/ptodata/2/ina/6_COMB.seq.*
6: /cgn2_6/ptodata/2/ina/6CTUS_COMB.seq.*
7: /cgn2_6/ptodata/2/ina/backfiles1.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	45.6	4.1	7218	1	US-08-232-463-14
C 2	37.4	3.4	6727	5	US-08-629-643A-5
C 3	37.2	3.4	2757	3	US-08-627-254C-26
C 4	36.6	3.3	5687	3	US-08-380-403A-3
C 5	36.6	3.3	5687	4	US-08-895-628-3
C 6	35.6	3.2	2908	5	US-08-487-799-1
C 7	35	3.2	9515	2	US-08-920-812-13
C 8	35	3.2	9515	2	US-08-920-827-13
C 9	35	3.2	9515	2	US-08-921-177-13
C 10	35	3.2	9515	2	US-08-362-577C-13
C 11	35	3.2	9515	2	US-08-920-828-13
C 12	34.8	3.2	17041	1	US-08-076-011-1
C 13	34.8	3.2	246240	3	US-08-724-394A-20
C 14	34.8	3.2	246240	3	US-08-724-394A-21
C 15	34.8	3.2	246240	3	US-08-724-394A-22
C 16	34.2	3.1	12847	3	US-08-550-715-1
C 17	33.6	3.1	3323	4	US-08-422-699A-10
C 18	33.6	3.1	3323	4	US-08-422-706B-10
C 19	33.4	3.0	35060	5	US-08-814-095-7
C 20	32.8	3.0	1334	4	US-08-481-658B-44
C 21	32.8	3.0	1334	4	US-08-477-504A-44
C 22	32.8	3.0	1334	4	US-08-486-756A-44
C 23	32.8	3.0	1334	4	US-08-485-862B-44
C 24	32.8	3.0	1334	5	US-08-787-739-44
C 25	32.8	3.0	6306	1	US-08-466-390-3
C 26	32.8	3.0	6306	2	US-08-470-950-3
C 27	32.8	3.0	6306	2	US-08-467-781-3

28	32.8	3.0	6306	2	US-08-195-487-3	Sequence 3, Appli
29	32.8	3.0	6306	3	US-08-483-924-3	Sequence 3, Appli
30	32.8	3.0	6306	6	PCT-US93-06160-3	Sequence 3, Appli
C 31	32.8	3.0	10898	4	US-08-481-658B-5	Sequence 5, Appli
C 32	32.8	3.0	10898	4	US-08-477-504A-5	Sequence 5, Appli
C 33	32.8	3.0	10898	4	US-08-486-756A-5	Sequence 5, Appli
C 34	32.8	3.0	10898	4	US-08-485-862B-5	Sequence 5, Appli
C 35	32.8	3.0	10898	5	US-08-787-739-5	Sequence 5, Appli
C 36	32.6	3.0	1294	4	US-08-665-647-4	Sequence 4, Appli
37	32.4	2.9	3957	1	US-07-689-008-5	Sequence 5, Appli
38	32.4	2.9	9540	1	US-07-689-008-1	Sequence 1, Appli
39	32	2.9	43795	5	US-08-742-185-101	Sequence 101, App
C 40	31.6	2.9	1548	5	US-08-665-259-28	Sequence 28, Appl
C 41	31.6	2.9	1589	1	US-07-971-092-1	Sequence 1, Appli
C 42	31.6	2.9	7676	2	US-08-451-777A-7	Sequence 7, Appli
C 43	31.6	2.9	7676	3	US-08-451-778A-7	Sequence 7, Appli
C 44	31.6	2.9	7676	3	US-08-998-208-7	Sequence 7, Appli
C 45	31.6	2.9	7676	6	PCT-US95-06743-7	Sequence 7, Appli

ALIGNMENTS

RESULT 1
US-08-232-463-14/c
; Sequence 14, Application US/08232463
; Patent No. 5670367
; GENERAL INFORMATION:
; APPLICANT: DORNER, F.
; APPLICANT: SCHEIFLINGER, F.
; APPLICANT: FALKNER, F. G.
; TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS
; NUMBER OF SEQUENCES: 52
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 1800 Diagonal Road, Suite 500
; CITY: Alexandria
; STATE: VA
; COUNTRY: USA
; ZIP: 22313-0299
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC Compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/232,463
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/07/935,313
; FILING DATE:
; APPLICATION NUMBER: EP 91 114 300.6
; FILING DATE: 26-AUG-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: BENT, Stephen A.
; REGISTRATION NUMBER: 29,768
; REFERENCE/DOCKET NUMBER: 30472/114 IMMU
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703)836-9300
; TELEFAX: (703)683-4109
; TELEX: 899149
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 7218 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; CLONE: pTZ9pt-Fls
; US-08-232-463-14

Query Match 4.1%; Score 45.6; DB 1; Length 7218;
Best Local Similarity 1.7%; Pred. No. 0.00087;
Matches 6; Conservative 207; Mismatches 141; Indels 0; Gaps 0;
QY 747 ttggcttaaggagatggcagaactggagaagtgagaagaagtgagcgaactgcaggactga 806
DB 1441 TTGGTACRR 1382
QY 807 ctgaacctggctactgtgggtgctgaagctggtacacacacacacaccccccactgtgat 866
DB 1381 RR 1322
QY 867 gagcccaactcattgaggtctcgtcatgtgagaaagctatttttaagtgaaagacagcggg 926
DB 1321 RR 1262
QY 927 actttcaggtttgtttatagtcacacagctggcagggcaggttataatctca 986
DB 1261 RR 1202
QY 987 ggccttgaagctgaggtgagaaatgggaagtgtaagctgggctgcttcattagtg 1046
DB 1201 RR 1142
QY 1047 aggcctcagtgctgaattaaaggaggtaagcaactattataaaaaaa 1100
DB 1141 RR 1088

RESULT 2
US-08-629-643A-5
; Sequence 5, Application US/08629643A
; Patent No. 6025539
; GENERAL INFORMATION:
; APPLICANT: Lee, J. L.
; APPLICANT: Lee, N. A.
; TITLE OF INVENTION: IL-5 TRANSGENIC MOUSE
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Schwegman, Lundberg, Woessner & Kluth, P. A.
; STREET: P. O. Box 2938
; CITY: Minneapolis
; STATE: MN
; COUNTRY: USA
; ZIP: 55402
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FASTSEQ Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/629,643A
; FILING DATE:
; CLASSIFICATION: 800
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/629,643
; FILING DATE: 09-APR-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Viksnins, Ann S
; REGISTRATION NUMBER: 37,748
; REFERENCE/DOCKET NUMBER: 150.167W01
; TELEPHONE: 612-359-3260
; TELEFAX: 612-359-3263
; TELEX:
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 6727 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: Genomic DNA
; FEATURE:

NAME/KEY: mat_peptide
LOCATION: 650...3771
OTHER INFORMATION: Join 650..730, 1560..1592,
OTHER INFORMATION: 3468..3596, 3676..3771
US-08-629-643A-5
Query Match 3.4%; Score 37.4; DB 5; Length 6727;
Best Local Similarity 50.9%; Pred. No. 0.22;
Matches 89; Conservative 0; Mismatches 86; Indels 0; Gaps 0;
QY 921 agcgggacttcaggtttgttttatagtcacacagctggcaggggtggcacagtttata 980
DB 1844 AGGTAGATGAAAGTTGTGCTTATAATAAAAGAAATATGAGAGTGGCAAAATATGTA 1903
QY 981 atctcagcccttgaagctgaggtggaagtgggaagtgaagctggcctggcttc 1040
DB 1904 ATCCAGTACTTGGGACCAAGCAGGGGTAGTCTGAGTCTAGGGCCAGCTTAGATACA 1963
QY 1041 atagtggctcagtgctgaattaaaggaggtaagcaactattataaaaaaa 1095
DB 1964 TTGCCCTGTATGATCAAAAGTAATCCTATAATAATAAACAACAAACATTAGA 2018

RESULT 3
US-08-627-254C-26
; Sequence 26, Application US/08627254C
; Patent No. 5859229
; GENERAL INFORMATION:
; APPLICANT: Kniss, Douglas A.
; TITLE OF INVENTION: Eicosanoid Formation
; NUMBER OF SEQUENCES: 29
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Calfee, Halter & Griwold LLP
; STREET: 800 Superior Avenue
; CITY: Cleveland
; STATE: Ohio
; COUNTRY: USA
; ZIP: 44114
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC Compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/627,254C
; FILING DATE:
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: Golrick, Mary E
; REGISTRATION NUMBER: 34,829
; REFERENCE/DOCKET NUMBER: 18525/00107
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (216) 622-8200
; TELEFAX: (216) 241-0816
; INFORMATION FOR SEQ ID NO: 126:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2757 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA to mRNA
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
US-08-627-254C-26
Query Match 3.4%; Score 37.2; DB 3; Length 2757;
Best Local Similarity 54.5%; Pred. No. 0.17;
Matches 97; Conservative 0; Mismatches 78; Indels 3; Gaps 1;
QY 925 ggacttcaggtttgttttatagtcacacagctggcaggggtggcacagtttataact 984

Db	2110	GGNATATAGGCTTAACCTTATATATATAGGGTAGGTGCTGGTTGCACACACACTTAATCC	21169
QY	985	cagcccttggaagctctgagcc---tggagaatggaagtgtaagctgggcctggccttca	1041
Db	2170	CAGCACTTGGAGGAGGAGGCGAGTTGGATCTCTGGAGCTTTGGCGGCAGTTTGGCCCTATA	2229
QY	1042	tagtagggctcagtgctcgaattaaagaggtgaagcaactattaaataaaaaaataaaaa	1099
Db	2230	TAGTAGGTTCTTAGGCCAGCCATGGATGCATAGTCAGACTCTTTCTCAAAACAACAAA	2287

RESULT 4
 US-08-380-403A-3/c
 : Sequence 3, Application US/08380403A
 : Patent No. 5831024
 :
 : GENERAL INFORMATION:
 :
 : APPLICANT: MINATO, Nagahiro
 : APPLICANT: HATORI, Masakazu
 : APPLICANT: HIROSHI, Kubota
 : APPLICANT: MASATSUGU, Masda
 :
 : TITLE OF INVENTION: SpA-1 PROTEIN AND GENE CODING THEREFOR
 :
 : NUMBER OF SEQUENCES: 7
 :
 : CORRESPONDENCE ADDRESS:
 :

Query Match	3.3%	Score 36.6;	DB 3;	Length 5687;
Best Local Similarity	56.1%;	Pred. No. 0.35;		
Matches 69;	Conservative	0;	Mismatches 54;	Indels 0;
Gaps 0;				

[illegible]

Qy	1037	tttcatagtgaggctcagtgctcgaattaaagaggtaaagcaactattataaaaaaaaaa	1099
Db <td>2552</td> <td>CTACAGAGTGTCTCCAGGATACCCAGGGCTACACAGAGAACCCCTGTCTTGAATAACA</td> <td>2593</td>	2552	CTACAGAGTGTCTCCAGGATACCCAGGGCTACACAGAGAACCCCTGTCTTGAATAACA	2593
Qy <td>1097</td> <td>aaa</td> <td>1099</td>	1097	aaa	1099
Db <td>2592</td> <td>AAA</td> <td>2590</td>	2592	AAA	2590

RESULT 5
 US-08-895-628-3/c
 Sequence 3, Application US/08895628
 Patent No. 5998585
 GENERAL INFORMATION:
 APPLICANT: MINATO, Nagahiro
 APPLICANT: HATTORI, Masakazu
 APPLICANT: HIROSHI, Kubota
 APPLICANT: MASATSUGU, Maeda
 TITLE OF INVENTION: SPA-1 PROTEIN AND GENE CODING THEREFOR
 NUMBER OF SEQUENCES: 7
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Foley & Lardner
 STREET: 3000 K Street, N.W., Suite 500
 CITY: Washington
 STATE: D.C.
 COUNTRY: USA
 ZIP: 20007-5109
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patent Release #1.0, Version #1.30
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/895,628
 FILING DATE:
 CLASSIFICATION: 536
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 08/380,403
 FILING DATE: 30-JAN-1995
 APPLICATION NUMBER: US 08/325,909
 FILING DATE: 19-OCT-1994
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: JP 6-279712
 FILING DATE: 20-OCT-1994
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: JP 6-139513
 FILING DATE: 30-MAY-1994
 ATTORNEY/AGENT INFORMATION:
 NAME: WEGNER, Harold C.
 REGISTRATION NUMBER: 25,258
 REFERENCE/DOCKET NUMBER: 53466/128/AAOK
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (202)672-5300
 TELEFAX: (202)672-5399
 TELEX: 904136
 INFORMATION FOR SEQ ID NO: 3:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 5687 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: double
 TOPOLOGY: linear
 MOLECULE TYPE: DNA (genomic)
 US-08-895-628-3

Query: Match	3.3%	Score 36.6;	DB 4;	Length 5687;
Best Local Similarity	56.1%	Pred. NO. 0.35;		
Matches 69;	Conservative	0;	Mismatches 54;	Indels 0;
				Gaps 0;

[illegible]

```
QY 1037 tttcatagtgaggctcagtgctgaattaaagaggttaagcaactattaaaaaa 1096
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2652 CTACAGAGTGAGTTCAGGATAGCCAGGGCTACACAGAAACCCGTGCTTGAAAAACAA 2593
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1097 aaa 1099
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2592 AAA 2590
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 6
US-08-487-799-1
; Sequence 1, Application US/08487799C
; Patent No. 6010908
; GENERAL INFORMATION:
; APPLICANT: Gruenert, Deiter C.
; TITLE OF INVENTION: GENE THERAPY BY SMALL FRAGMENTS HOMOLOGOUS REPLACEMENT
; FILE REFERENCE: 480.18-1(HV)
; CURRENT APPLICATION NUMBER: US/08/487,799C
; CURRENT FILING DATE: 1995-06-07
; EARLIER APPLICATION NUMBER: 07/933,471
; EARLIER FILING DATE: 1992-08-21
; EARLIER APPLICATION NUMBER: 08/409,544
; EARLIER FILING DATE: 1995-03-24
; NUMBER OF SEQ ID NOS: 87
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO 1
; LENGTH: 2908
; TYPE: DNA
; ORGANISM: human
US-08-487-799-1

Query Match 3.2%; Score 35.6; DB 5; Length 2908;
Best Local Similarity 71.2%; Pred. No. 0.51;
Matches 47; Conservative 0; Mismatches 19; Indels 0; Gaps 0;

QY 1030 gctggtcttcctagtgagctcagtgctgaattaaagaggttaagcaactattaaaaaa 1089
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 126 gctggtggtgacagtgagtgactctgtcttaataataataataataataataataa 185
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1090 aaaaaa 1095
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 186 ataaaa 191
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 7
US-08-920-812-13/c
; Sequence 13, Application US/08920812
; Patent No. 5763188
; GENERAL INFORMATION:
; APPLICANT: Ohno, Tsuneya
; APPLICANT: Matsuhisa, Akio
; APPLICANT: Uehara, Hirotsugu
; APPLICANT: Eda, Soji
; TITLE OF INVENTION: Probe for Diagnosing Infectious Disease
; NUMBER OF SEQUENCES: 25
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
; STREET: 6300 Sears Tower, 233 South Wacker Drive
; CITY: Chicago
; STATE: Illinois
; COUNTRY: United States of America
; ZIP: 60606-6402
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/920,812
; FILING DATE: 29-AUG-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/362,577
; FILING DATE: 27-MAR-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Rin-Laures, Li-Hsien
; REGISTRATION NUMBER: 33,547
; REFERENCE/DOCKET NUMBER: 19036/32420
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 312/474-6300
```

```
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/362,577
; FILING DATE: 27-MAR-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Rin-Laures, Li-Hsien
; REGISTRATION NUMBER: 33,547
; REFERENCE/DOCKET NUMBER: 19036/32420
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 312/474-6300
; TELEFAX: 312/474-0448
; TELEX: 25-3856
; INFORMATION FOR SEQ ID NO: 13:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 9515 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: Genomic DNA
; ORIGINAL SOURCE:
; ORGANISM: Pseudomonas aeruginosa
; STRAIN: Clinical Isolate P2-2
US-08-920-812-13

Query Match 3.2%; Score 35; DB 2; Length 9515;
Best Local Similarity 56.5%; Pred. No. 1.3;
Matches 65; Conservative 0; Mismatches 50; Indels 0; Gaps 0;

QY 318 gactgcagagaagctgtggagcaacttggtgcagcgccaaacctgtcagaagagcgtctg 377
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1222 GACCTCGGTGGAGCGGGTGAACACACAGTCGCTCCACCTCCGGGTGCAAGCGAGGGCCTT 1163
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 378 gccgtcctctggtggcgccacacacacctgctccagcaggtctccggtctccccc 432
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1162 GCCGCGCTGCTCGCGAGGCGCGGCACACACAGCTTACGACACGCTTCCGGCACCCCC 1108
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 8
US-08-920-827-13/c
; Sequence 13, Application US/08920827
; Patent No. 5770375
; GENERAL INFORMATION:
; APPLICANT: Ohno, Tsuneya
; APPLICANT: Matsuhisa, Akio
; APPLICANT: Uehara, Hirotsugu
; APPLICANT: Eda, Soji
; TITLE OF INVENTION: Probe for Diagnosing Infectious Disease
; NUMBER OF SEQUENCES: 25
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
; STREET: 6300 Sears Tower, 233 South Wacker Drive
; CITY: Chicago
; STATE: Illinois
; COUNTRY: United States of America
; ZIP: 60606-6402
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/920,827
; FILING DATE: 29-AUG-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/362,577
; FILING DATE: 27-MAR-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Rin-Laures, Li-Hsien
; REGISTRATION NUMBER: 33,547
; REFERENCE/DOCKET NUMBER: 19036/32420
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 312/474-6300
```



```
;
; TELEFAX: 312/474-0448
; TELEX: 25-3856
; INFORMATION FOR SEQ ID NO: 13:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 9515 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: Genomic DNA
; ORIGINAL SOURCE:
; ORGANISM: Pseudomonas aeruginosa
; STRAIN: Clinical isolate P2-2
;
US-08-920-827-13

Query Match 3.2%; Score 35; DB 2; Length 9515;
Best Local Similarity 56.5%; Pred. No. 1.3;
Matches 65; Conservative 0; Mismatches 50; Indels 0; Gaps 0;

QY 318 gactgcagagaagctgtggagcaacttggtgccagcgcccaacctgtgcagaagagcgtctg 377
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1222 GACCTCGGTGGAGCGCGTGAACACCACTGCGTCCACCTCGGGTGCAAGCCGAGGGCCTT 1163

QY 378 gccctcctctggtggggcacacacacacctgtccagcaggtctccgggtgcccc 432
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1162 GCCGGCTGCTCGCCGAGCGCCGACCACTGTCAGCACGCTTCCGGGCACCCCC 1108

RESULT 9
US-08-921-177-13/C
; Sequence 13, Application US/08921177
; Patent No. 5798211
; GENERAL INFORMATION:
; APPLICANT: Ohno, Tsuneya
; APPLICANT: Matsuhisa, Akio
; APPLICANT: Uehara, Hirotugu
; APPLICANT: Eda, Soji
; TITLE OF INVENTION: Probe for Diagnosing Infectious Disease
; NUMBER OF SEQUENCES: 25
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
; STREET: 6300 Sears Tower, 233 South Wacker Drive
; CITY: Chicago
; STATE: Illinois
; COUNTRY: United States of America
; ZIP: 60606-6402
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/921,177
; FILING DATE: 29-AUG-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/362,577
; FILING DATE: 27-MAR-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Rin-Laures, Li-Hsien
; REGISTRATION NUMBER: 33,547
; REFERENCE/DOCKET NUMBER: 19036/32420
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 312/474-6300
; TELEFAX: 312/474-0448
; TELEX: 25-3856
; INFORMATION FOR SEQ ID NO: 13:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 9515 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: Genomic DNA
; ORIGINAL SOURCE:
; ORGANISM: Pseudomonas aeruginosa
; STRAIN: Clinical isolate P2-2
;
US-08-920-827-13

Query Match 3.2%; Score 35; DB 2; Length 9515;
Best Local Similarity 56.5%; Pred. No. 1.3;
Matches 65; Conservative 0; Mismatches 50; Indels 0; Gaps 0;

QY 318 gactgcagagaagctgtggagcaacttggtgccagcgcccaacctgtgcagaagagcgtctg 377
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1222 GACCTCGGTGGAGCGCGTGAACACCACTGCGTCCACCTCGGGTGCAAGCCGAGGGCCTT 1163

QY 378 gccctcctctggtggggcacacacacacctgtccagcaggtctccgggtgcccc 432
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1162 GCCGGCTGCTCGCCGAGCGCCGACCACTGTCAGCACGCTTCCGGGCACCCCC 1108

RESULT 9
US-08-921-177-13/C
; Sequence 13, Application US/08921177
; Patent No. 5798211
; GENERAL INFORMATION:
; APPLICANT: Ohno, Tsuneya
; APPLICANT: Matsuhisa, Akio
; APPLICANT: Uehara, Hirotugu
; APPLICANT: Eda, Soji
; TITLE OF INVENTION: Probe for Diagnosing Infectious Disease
; NUMBER OF SEQUENCES: 25
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
; STREET: 6300 Sears Tower, 233 South Wacker Drive
; CITY: Chicago
; STATE: Illinois
; COUNTRY: United States of America
; ZIP: 60606-6402
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/921,177
; FILING DATE: 29-AUG-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/362,577
; FILING DATE: 27-MAR-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Rin-Laures, Li-Hsien
; REGISTRATION NUMBER: 33,547
; REFERENCE/DOCKET NUMBER: 19036/32420
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 312/474-6300
; TELEFAX: 312/474-0448
; TELEX: 25-3856
; INFORMATION FOR SEQ ID NO: 13:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 9515 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: Genomic DNA
; ORIGINAL SOURCE:
; ORGANISM: Pseudomonas aeruginosa
; STRAIN: Clinical isolate P2-2
;
US-08-920-827-13
```

```
;
; ORIGINAL SOURCE:
; ORGANISM: Pseudomonas aeruginosa
; STRAIN: Clinical isolate P2-2
;
US-08-921-177-13

Query Match 3.2%; Score 35; DB 2; Length 9515;
Best Local Similarity 56.5%; Pred. No. 1.3;
Matches 65; Conservative 0; Mismatches 50; Indels 0; Gaps 0;

QY 318 gactgcagagaagctgtggagcaacttggtgccagcgcccaacctgtgcagaagagcgtctg 377
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1222 GACCTCGGTGGAGCGCGTGAACACCACTGCGTCCACCTCGGGTGCAAGCCGAGGGCCTT 1163

QY 378 gccctcctctggtggggcacacacacacctgtccagcaggtctccgggtgcccc 432
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1162 GCCGGCTGCTCGCCGAGCGCCGACCACTGTCAGCACGCTTCCGGGCACCCCC 1108

RESULT 10
US-08-362-577C-13/C
; Sequence 13, Application US/08362577C
; Patent No. 5807673
; GENERAL INFORMATION:
; APPLICANT: Ohno, Tsuneya
; APPLICANT: Matsuhisa, Akio
; APPLICANT: Uehara, Hirotugu
; APPLICANT: Eda, Soji
; TITLE OF INVENTION: Probe for Diagnosing Infectious Disease
; NUMBER OF SEQUENCES: 25
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
; STREET: 6300 Sears Tower, 233 South Wacker Drive
; CITY: Chicago
; STATE: Illinois
; COUNTRY: United States of America
; ZIP: 60606-6402
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/362,577C
; FILING DATE: 27-MAR-1995
; CLASSIFICATION: 536
; ATTORNEY/AGENT INFORMATION:
; NAME: Rin-Laures, Li-Hsien
; REGISTRATION NUMBER: 33,547
; REFERENCE/DOCKET NUMBER: 19036/32420
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 312/474-6300
; TELEFAX: 312/474-0448
; TELEX: 25-3856
; INFORMATION FOR SEQ ID NO: 13:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 9515 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: Genomic DNA
; ORIGINAL SOURCE:
; ORGANISM: Pseudomonas aeruginosa
; STRAIN: Clinical isolate P2-2
;
US-08-362-577C-13

Query Match 3.2%; Score 35; DB 2; Length 9515;
Best Local Similarity 56.5%; Pred. No. 1.3;
Matches 65; Conservative 0; Mismatches 50; Indels 0; Gaps 0;

QY 318 gactgcagagaagctgtggagcaacttggtgccagcgcccaacctgtgcagaagagcgtctg 377
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```

Db 1222 GACCTCGTGGAGCGGTGAACACACAGTGCCTCCACCTCCGGGTGCAAGCGAGGGCCTT 1163

Qy 378 gccgtctgctggcgggcacacacaccctgtccagcagctctccggctcccc 432

Db 1162 GCCGGCCTGCTCGCGGAGGCCCGGACCAACCGTTACGACGCTTCCGGCACCCCC 1108

RESULT 11

US-08-920-828-13/C
; Sequence 13, Application US/08920828
; Patent No. 5853998
; GENERAL INFORMATION:
; APPLICANT: Ohno, Tsuneya
; APPLICANT: Matsuhisa, Akio
; APPLICANT: Uehara, Hirotugu
; APPLICANT: Eda, Soji
; TITLE OF INVENTION: Probe for Diagnosing Infectious Disease
; NUMBER OF SEQUENCES: 25
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
; STREET: 6300 Sears Tower, 233 South Wacker Drive
; CITY: Chicago
; STATE: Illinois
; COUNTRY: United States of America
; ZIP: 60606-6402

COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/920,828
; FILING DATE: 29-AUG-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/362,577
; FILING DATE: 27-MAR-1995

ATTORNEY/AGENT INFORMATION:
; NAME: Rin-Laures, Li-Hsien
; REGISTRATION NUMBER: 33,547
; REFERENCE/DOCKET NUMBER: 19036/32420
; TELEPHONE: 312/474-6500
; TELEFAX: 312/474-0448
; TELEX: 25-3856

INFORMATION FOR SEQ ID NO: 13:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 9515 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: Genomic DNA
; ORIGINAL SOURCE:
; ORGANISM: Pseudomonas aeruginosa
; STRAIN: Clinical Isolate P2-2

US-08-920-828-13
Query Match 3.2%; Score 35; DB 3; Length 9515;
Best Local Similarity 56.5%; Pred. No. 1.3;
Matches 65; Conservative 0; Mismatches 50; Indels 0; Gaps 0;

Qy 318 gactgcagagaagctgtgagcaacttggccagcgccaacctgtcagaagagcgtctg 377

Db 1222 GACCTCGTGGAGCGGTGAACACACAGTGCCTCCACCTCCGGGTGCAAGCGAGGGCCTT 1163

Qy 378 gccgtctgctggcgggcacacacaccctgtctccagcagctctccggctgcccc 432

Db 1162 GCCGGCCTGCTCGCGGAGGCCCGGACCAACCGTTACGACGCTTCCGGCACCCCC 1108

RESULT 12

US-08-076-011-1/C

; Sequence 1, Application US/08076011
; Patent No. 5521069
; GENERAL INFORMATION:
; APPLICANT: ONDA, Haruo
; APPLICANT: KIMURA, Chiharu
; APPLICANT: OHKUBO, Shoichi
; TITLE OF INVENTION: NOVEL DNA AND USE THEREOF
; NUMBER OF SEQUENCES: 3
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: DAVID G. CONLIN; DIKE, BRONSTEIN,
; ADDRESSEE: ROBERTS & CUSHMAN
; STREET: 130 Water Street
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: US
; ZIP: 02109

COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/076,011
; FILING DATE: 11-JUN-1993
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/047,246
; FILING DATE: 13-APR-1993

NAME:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/741,676
; FILING DATE: 07-AUG-1991

ATTORNEY/AGENT INFORMATION:
; NAME: RESNICK, DAVID S
; REGISTRATION NUMBER: 34235
; REFERENCE/DOCKET NUMBER: 41155-CIP
; TELEPHONE: (617)523-3400
; TELEFAX: (617)523-6440
; TELEX: 200291 STRE UR

INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 17041 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)

FEATURE:
; NAME/KEY: CDS
; LOCATION: join(7540..7650, 9814..9945, 10421..10519,
; LOCATION: 11602..11787)
US-08-076-011-1

Query Match 3.2%; Score 34.8; DB 1; Length 17041;
Best Local Similarity 60.6%; Pred. No. 2;
Matches 57; Conservative 0; Mismatches 37; Indels 0; Gaps 0;

Qy 989 ccttggaaagctgagcgtggaatggaagctgaaagctggcctgttcacatgtgag 1048

Db 14797 CTTTGGGAGGCTGAGTGGGAGGATCGCTTGAGCAAGACCAGCTGGCAACATAGTGA 14738

Qy 1049 gctcagtcgcaattaaagaggtaaagcaactat 1082

Db 14737 ACCTGCCACTAAATAAAAAATAAAGACAAT 14704

RESULT 13

US-08-724-394A-20
; Sequence 20, Application US/08724394A
; Patent No. 5872237
; GENERAL INFORMATION:

```

APPLICANT: Feder, John N.
APPLICANT: Kronmal, Gregory S.
APPLICANT: Lauer, Peter M.
APPLICANT: Ruddy, David A.
APPLICANT: Thomas, Winston
APPLICANT: Tsuchihashi, Zenta
APPLICANT: Wolff, Roger K.
TITLE OF INVENTION: Megabase Transcript Map: No. 5872237el
TITLE OF INVENTION: Sequences and Antibodies Thereeto
NUMBER OF SEQUENCES: 31
CORRESPONDENCE ADDRESS:
ADDRESSEE: TOWNSEND and TOWNSEND and CREW LLP
STREET: Two Embarcadero Center, 8th Floor
CITY: San Francisco
STATE: CA
COUNTRY: USA
ZIP: 94111-3834
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/724.394A
FILING DATE: 01-OCT-1996
CLASSIFICATION: 536
ATTORNEY/AGENT INFORMATION:
NAME: Fitts, Renee A.
REGISTRATION NUMBER: 35,136
REFERENCE/DOCKET NUMBER: 017957-000100
TELEPHONE: 415-576-0200
TELEFAX: 415-576-0300
INFORMATION FOR SEQ ID NO: 20:
SEQUENCE CHARACTERISTICS:
LENGTH: 246240 base pairs
TYPE: nucleic acid
STRANDEDNESS: not relevant
TOPOLOGY: not relevant
MOLECULE TYPE: cDNA
FEATURE:
NAME/KEY: misc.feature
LOCATION: 1..246240
OTHER INFORMATION: /note= "HLA-H.CONTIG"
US-08-724-394A-20

Query Match          3.2%; Score 34.8; DB 3; Length 246240;
Best Local Similarity 57.3%; Pred. No. 7;
Matches 63; Conservative 0; Mismatches 47; Indels 0; Gaps 0

QY    990   ctctggaggtcgaggctggaagaatggaaagttaaacttataaaaacacaaaaaaagtaaacgcatttccttagtgagg 1049
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 233301 CTGGGAGGCTGAGCGAAGAATCACTTGAATATAGGAGGGGAGGTTCACAGTGAGAGT 233360

QY    1050   ctcaagtctgaattaagaagtgtaaagcaacctattataaaaaaaaaaaaaa 1099
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 233361 GAGACTCTCAAAAAAAACAACAACAAAAGTAACAAGAAAAAGAAA 233410


RESULT 14
US-08-724-394A-21
Sequence 21, Application US/08724394A
Patent No. 5872237
GENERAL INFORMATION:
APPLICANT: Feder, John N.
APPLICANT: Kronmal, Gregory S.
APPLICANT: Lauer, Peter M.
APPLICANT: Ruddy, David A.
APPLICANT: Thomas, Winston
APPLICANT: Tsuchihashi, Zenta
APPLICANT: Wolff, Roger K.
TITLE OF INVENTION: Megabase Transcript Map: No. 5872237el
TITLE OF INVENTION: Sequences and Antibodies thereco
NUMBER OF SEQUENCES: 31
CORRESPONDENCE ADDRESS:
ADDRESSEE: TOWNSEND and TOWNSEND and CREW LLP
STREET: Two Embarcadero Center, 8th Floor
CITY: San Francisco
STATE: CA
COUNTRY: USA

APPLICANT: Feder, John N.
APPLICANT: Kronmal, Gregory S.
APPLICANT: Lauer, Peter M.
APPLICANT: Ruddy, David A.
APPLICANT: Thomas, Winston
APPLICANT: Tsuchihashi, Zenta
APPLICANT: Wolff, Roger K.
TITLE OF INVENTION: Megabase Transcript Map: No. 5872237el
TITLE OF INVENTION: Sequences and Antibodies Therto
NUMBER OF SEQUENCES: 31
CORRESPONDENCE ADDRESS:
ADDRESSEE: TOWNSEND and TOWNSEND and CREW LLP
STREET: Two Embarcadero Center, 8th Floor
CITY: San Francisco
STATE: CA
COUNTRY: USA
ZIP: 94111-3834
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/724.394A
FILING DATE: 01-OCT-1996
CLASSIFICATION: 536
ATTORNEY/AGENT INFORMATION:
NAME: Fitts, Renee A.
REGISTRATION NUMBER: 35,136
REFERENCE/DOCKET NUMBER: 017957-000100
TELEPHONE: 415-576-0200
TELEFAX: 415-576-0300
INFORMATION FOR SEQ ID NO: 20:
SEQUENCE CHARACTERISTICS:
LENGTH: 246240 base pairs
TYPE: nucleic acid
STRANDEDNESS: not relevant
TOPOLOGY: not relevant
MOLECULE TYPE: cDNA
FEATURE:
NAME/KEY: misc.feature
LOCATION: 1..246240
OTHER INFORMATION: /note= "HLA-H.CONTIG"
US-08-724-394A-20

Query Match          3.2%; Score 34.8; DB 3; Length 246240;
Best Local Similarity 57.3%; Pred. No. 7;
Matches 63; Conservative 0; Mismatches 47; Indels 0; Gaps 0;

QY    990   ctctggaggtcgaggctggaagaatggaaagttaaacttataaaaacacaaaaaaagtaaacgcatttccttagtgagg 1049
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 233301 CTGGGAGGCTGAGCGAAGAATCACTTGAATATAGGAGGGGAGGTTCACAGTGAGAGT 233360

QY    1050   ctcaagtctgaattaagaagtgtaaagcaacctattataaaaacacaaaaaaagtaaacgcatttccttagtgagg 1099
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 233361 GAGACTCTCAAAAAAAACAACAACAAAAGTAACAAGAAAAAGAAA 233410


RESULT 14
US-08-724-394A-21
Sequence 21, Application US/08724394A
Patent No. 5872237
GENERAL INFORMATION:
APPLICANT: Feder, John N.
APPLICANT: Kronmal, Gregory S.
APPLICANT: Lauer, Peter M.
APPLICANT: Ruddy, David A.
APPLICANT: Thomas, Winston
APPLICANT: Tsuchihashi, Zenta
APPLICANT: Wolff, Roger K.
TITLE OF INVENTION: Megabase Transcript Map: No. 5872237el
TITLE OF INVENTION: Sequences and Antibodies Therto
NUMBER OF SEQUENCES: 31
CORRESPONDENCE ADDRESS:
ADDRESSEE: TOWNSEND and TOWNSEND and CREW LLP
STREET: Two Embarcadero Center, 8th Floor
CITY: San Francisco
STATE: CA
COUNTRY: USA

```

```

; ZIP: 94111-3834
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/724,394A
; FILING DATE: 01-OCT-1996
; CLASSIFICATION: 536
; ATTORNEY/AGENT INFORMATION:
; NAME: Fitts, Renee A.
; REGISTRATION NUMBER: 35,136
; REFERENCE/DOCKET NUMBER: 017957-000100
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-576-0200
; TELEFAX: 415-576-0300
; INFORMATION FOR SEQ ID NO: 22:
; LENGTH: 246240 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: not relevant
; TOPOLOGY: not relevant
; MOLECULE TYPE: cdna
; FEATURE:
; NAME/KEY: misc.feature
; LOCATION: 1..246240
; OTHER INFORMATION: /note= "HLA-H.CONTIG"
US-08-724-394A-22

```

```

Query Match          3.2%; Score 34.8; DB 3; Length 246240;
Best Local Similarity 57.3%; Pred No. 7;
Matches 63; Conservative 0; Mismatches 47; Indels 0; Gaps 0;

QY 990 cttggaagtctgaggctggagaagtgaagctggcctgcttcacatagtggagg 1049
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 233301 CTCGGGAGGCTGAGGCAAGAGAACTTGAATATAGGAGGGGGAGGTTGCAGTGAGACT 233360

QY 1050 ctcaagtctgaattaaaggaggtaaaagcaactattaaaaaagagagagagag 1099
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 233361 GAGACTCTCAAAAAAACAACAAAAAAGTAACAAAGAAAAAGAAA 233410

```

Search completed: May 15, 2000, 01:50:08
Job time: 10796 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 14, 2000, 23:02:16 ; Search time 84.19 Seconds
(without alignments)
3268.931 Million cell updates/sec

Title: US-09-223-796-1

Perfect score: 1100

Sequence: 1 gcacgagccacagccagccta.....attaaaaaaaaaaaaaaaa 1100

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 311585 seqs, 125096042 residues

Total number of hits satisfying chosen parameters: 623170

Minimum DB seq length: 0

Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database : N_Geneseq_36.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1100	100.0	1100	1	V18890 Rat Hypertension r
2	149.2	13.6	701	1	V18891 Human Hypertension
3	41.4	3.8	218	1	T32035 Proliferation-inhi
4	39.8	3.6	849	1	V06127 Viral infection ge
5	37.4	3.4	6727	1	T88014 Murine IL-5 cDNA g
6	36.6	3.3	2858	1	Q75270 Unique 2.9 kb geno
7	36.6	3.3	5687	1	T65001 Mouse cell cycle r
8	36.4	3.3	4736	1	T23315 Mouse 1-alpha-OHs
9	36.2	3.3	602	1	X20419 Human secreted pro
10	36.2	3.3	1371	1	V33755 Mouse neutrophil e
11	36.2	3.3	2851	1	Q72975 Cancer repression
12	36.2	3.3	5150	1	T78598 AS-30D tumour type
13	36.2	3.3	14180	1	V05159 Mouse butyrophilin
14	35.8	3.3	921	1	X51728 DNA encoding a hum
15	35.4	3.2	1089	1	X37460 Human secreted pro
16	35.4	3.2	2880	1	X18193 hNBC3a coding sequ
17	35.4	3.2	3912	1	X18192 hNBC3 coding sequ
18	35	3.2	2458	1	T89399 Methods for diagno
19	35	3.2	8353	1	T89396 Methods for diagno
20	35	3.2	9515	1	Q55145 Pseudomonas aerugi
21	34.8	3.2	17041	1	Q21065 Genomic DNA of hum
22	34.8	3.2	17041	1	Q67057 PACAP38 DNA. PACA
23	34.8	3.2	237326	1	V57903 Hereditary haemoch
24	34.2	3.1	2560	1	X18941 Human U62317 gene
25	33.8	3.1	1573	1	V43607 Human secreted pro
26	33.8	3.1	4630	1	V84139 Human prothrombin
27	33.6	3.1	1698	1	X23318 Mouse 1-alpha-OHs
28	33.6	3.1	3323	1	Q48675 Human myotonic dys
29	33.6	3.1	4105	1	X23316 Mouse 1-alpha-OHs
30	33.4	3.0	2344	1	T89885 Drosophila frizzles
31	33	3.0	13206	1	Q05243 Rat acyl peptide h
32	33	3.0	50000	1	X23517 Human kidney amino
33	32.8	3.0	392	1	Q61201 Human brain Expres
34	32.8	3.0	6306	1	Q54841 Sequence encoding

35	32.8	3.0	6579	1	T77783	CDNA encoding Lexa
36	32.8	3.0	6819	1	T77782	CDNA encoding GAL4
37	32.8	3.0	10897	1	T09187	Muru putative onco
38	32.6	3.0	655	1	X37545	Human secreted pro
39	32.6	3.0	1294	1	T60587	Protein cognate of
40	32.6	3.0	3848	1	X27277	Prostate-tumour de
41	32.6	3.0	5668	1	X27262	Prostate-tumour de
42	32.6	3.0	5873	1	T61701	Human prostatic sp
43	32.4	2.9	3360	1	V19608	Mus musculus EPCR
44	32.4	2.9	10169	1	Q53522	Cellulose synthase
45	32	2.9	1908	1	V73002	Human adult brain

ALIGNMENTS

RESULT 1

V18890 Vi8890 standard; cDNA; 1100 BP.
AC V18890;
DE 09-JUN-1998 (first entry)
DE Rat Hypertension related calcium regulated gene (HcARG) cDNA.
KW Hypertension related calcium regulated gene; HcARG; rat parathyroid;
KW extracellular calcium concentration; antibody; hypertension; ss;
KW hyperthyroidism; osteoporosis; heart failure; diabetes; stroke;
KW cancer; inflammatory disease; asthma.
OS Rattus rattus.
PH Key Location/Qualifiers
FT CDS 131..806
FT /*tag= a
FT /*product= "Hypertension related calcium regulator"
FT misc_feature 173..196
FT /*tag= b
FT /*note= "EF-hand like motif"
PN W09749807-A2.
PD 31-DEC-1997.
PF 23-JUN-1997; CA0439.
PR 21-JUN-1996; US-667495.
PA (GOSS/) GOSSARD F.
PA (HAME/) HAMEY P.
PA (LEWA/) LEWANCZUK R.
PA (TREM/) TREMBLAY J.
PI Gossard F, Hamet P, Lewanczuk R, Tremblay J;
WPI; 98-077171/07.
DR P-PSDB; W37723.
DR Hypertension related calcium regulated gene - useful to develop
PT products to treat or detect, e.g. hypertension, stroke,
PT osteoporosis, heart failure, cancer, diabetes or asthma
PS Claim 2; Pages 25-26; 46pp; English.
CC This nucleic acid sequence encodes the hypertension related calcium
CC regulated gene (HcARG), which was isolated from the rat parathyroid.
CC Its expression is regulated by extracellular calcium concentration.
CC An antibody against the protein, can be used to detect or modulate
CC (e.g. enhance or inhibit) abnormal calcium levels. They can
CC specifically be used to detect or treat, e.g. hypertension,
CC hyperthyroidism, osteoporosis, heart failure, diabetes, cancer,
CC inflammatory disease, and asthma.
SQ Sequence 1100 BP; 260 A; 279 C; 315 G; 246 T;

Query Match 100.0%; Score 1100; DB 1; Length 1100;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1100; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	1	gcacgagccacagccagctaccgcgctaggttcctccaggtgcagagcggttaagc	60
DB	1	GCACGAGCCACAGCCAGCTACCGCGGCTAGTTCCTCCAGGTGCAGAGCGGCTAAAGC	60
QY	61	ttggtttgtattgtaactgacactggtagacctctcttcgactgttcagaagac	120
DB	61	TTGGTTGTATTGTATGTAATGCACTGGTTAGGACCTTCCTTCGGAGCTGTCAGAAAC	120
QY	121	gggaagaaagatgtctgcttgggggggctgcagctccactgtccaccatcccgctgaca	180

Db 121 GGGAAAGAGGATGTCGCTTTGGGGCTGCAGCTCCATACCTGACCAATCCGCTGACA 180
 Qy 181 gtcaagtaggcgggtcagttctctggtgtccagccctctcagaagtagcggcgtgg 240
 Db 181 GTCACAGTGGCGGGTCACTTCTGGGTTCAGAGCCCTCTCAGAAGTGCAGGGCGTGG 240
 Qy 241 ctcaagctctgaaggaacttagacagagacaccttcagaaagtgttgaaactgttagtgc 300
 Db 241 CTCAGCTCTTGAAGGACTTAGACAGAGACACCTTCAGAAAGTGTGAACTGTAGTGC 300
 Qy 301 gggccctgcattgggaagactgcagagaagctgtgagcaacttggccagcgcaacc 360
 Db 301 GGGCCCTGCATGGAAAGACTGCAGAGAAGCTGTGGAGCAACTTGTGCAGCGCAACC 360
 Qy 361 tgtcaagaagaggtctggcgtctctgctgagcgggacacacacccctgctccagcagctc 420
 Db 361 TGTCAAGAGAGCGTCTGGCGCTCTCTGCTGGCGGACACACACCCCTGCTCCAGCAGGCTC 420
 Qy 421 tccgggtgccccctgtagtctaaagccagatgccttcaggaagagctccaggaacttg 480
 Db 421 TCGGGTGTCCCTCTGTAGTCTAAAGCCAGATGCTTCCAGGAGAGCTCCAGGAACCTG 480
 Qy 481 gcaattcctcaggatctaattggagattggccagatttggcatttggagtagcaacgcctc 540
 Db 481 GCATTCTCAGGATCTAATTGGAGATTGGCCAGATTGGCAATTGGGAGTCAACGCCCTC 540
 Qy 541 ttctcgaactctgagcccaacagcaggaatcctcgtcctcagctcagctgttactccgt 600
 Db 541 TTCTCAGCTCTGTAGGCCCAACAGCAGGAGTCTCGCTGCTCAGTGTCTTACTTCCGT 600
 Qy 601 ggcgggtgagctggtgcattctcaacagcgtcagtcctcagcctccctgcacagagtgctc 660
 Db 601 GCGGGTGGAGTGGCCATCTCAACAGCGCTCAGTCCCGCTCCCTGCAACCGAGTCTC 660
 Qy 661 tcatgagctgaagctcaagatgagatgagctgacacccctgagctgagctgagctgagct 720
 Db 661 TCATGAGCTGAAGCTCACAGATGAGTCTGCACACCGCTTCGAGGTGCCCATAGCCAAAT 720
 Qy 721 ttcaagagctcggttagctgagctgtgctccttaagagagtgagagctgagagaga 780
 Db 721 TTCAAGAGCTGCGGTACAGTGTAGCTTGTGCTTAAAGAGATGGCAGAACTGGAGAGA 780
 Qy 781 agttagcgcgaactgcaggaactgactgaacccctggtactggtggtggtgagctggta 840
 Db 781 AGTGTGAGCGCAAACTGCAGGACTGACTGAACCTCTGTACTGTGGTGTGTAAGCTGGTA 840
 Qy 841 ccagaacacagccccccactggtgagcggcgaactccactcattgaggtcctgcatgtgaga 900
 Db 841 CCAGAACACAGCCCCCCTGATGATGAGCCCAACTCCATTGAGGTCTCTGCATGTGAGAA 900
 Qy 901 cgtattttagtgaagacagcggacttccaggtttgttttatgagtcacacagctgg 960
 Db 901 CGTATTTTAAAGTGAAGACAGCGGAGACTTTCAGGTTTGTGTTTATGAGTCAACACTGG 960
 Qy 961 gcaggtgtgacaggttataatctcagcccttggagagctgagctgagctggagagatgga 1020
 Db 961 GCAGGTGGCACAGTTTATAATCTCAGCCCTTGGAACTGTGAGGCTGGAGAAATGGGAAGT 1020
 Qy 1021 gtaagctggccttgcattcagctgaggtcagctgagctgagctgagctgagctgagct 1080
 Db 1021 GTAAGCTGGCCCTGGCTTTCATAGTGGGCTCAGTGTGCAATTTAAAGAGTAAAGCAACT 1080
 Qy 1081 attaaaaaataaaaaa 1100
 Db 1081 ATTAATAAAAAAAAAA 1100

RESULT 2
 ID V18891
 AC V18891;
 DT 09-JUN-1998 (first entry)

DE Human Hypertension related calcium regulated gene (HCAeG) cDNA.
 KW Hypertension related calcium regulated gene; HCAeG; human parathyroid;
 KW extracellular calcium concentration; antibody; hypertension; ss;
 KW hyperthyroidism; osteoporosis; heart failure; diabetes; stroke;
 OS cancer; inflammatory disease; asthma.
 FH Homo sapiens.
 Key Location/Qualifiers
 misc_feature 224..245
 /*tag= a
 /*note= "EF-hand like motif"
 W09749807-A2.
 PD 31-DEC-1997. CA0439.
 PF 23-JUN-1997; US-667495.
 PR 21-JUN-1996; US-667495.
 PA (GOSS/) GOSSARD F.
 PA (HAME/) HAME P.
 PA (LEWA/) LEWANCZUK R.
 PA (TREM/) TREMBLAY J.
 PI Gossard F, Hamet P, Lewanczuk R, Tremblay J;
 WPI: 98-077171/07.
 DR Hypertension related calcium regulated gene - useful to develop
 PT products to treat or detect, e.g. hypertension, stroke,
 PT osteoporosis, heart failure, cancer, diabetes or asthma
 PS Claim 3; Pages 27-28; 46pp; English.
 CC This nucleic acid sequence encodes the hypertension related calcium
 CC regulated gene (HCAeG), which was isolated from the human parathyroid.
 CC Its expression is regulated by extracellular calcium concentration.
 CC The sequence displays 75 percent homology to the rat form, with
 CC conserved areas such as the EF-hand like motifs and the initiating
 CC codon. An antibody against the protein, can be used to detect or
 CC modulate (e.g. enhance or inhibit) abnormal calcium levels. They can
 CC specifically be used to detect or treat, e.g. hypertension,
 CC hyperthyroidism, osteoporosis, heart failure, diabetes, stroke, cancer,
 CC inflammatory disease, and asthma.
 SQ Sequence 701 BP; 137 A; 204 C; 217 G; 139 T;

Query Match 13.6%; Score 149.2; DB 1; Length 701;
 Best Local Similarity 70.1%; Pred. No. 8.9e-33;
 Matches 213; Conservative 0; Mismatches 90; Indels 1; Gaps 1;
 Qy 88 gtagagacattctctcaggaactggtcagaagaacgggaagagatgctgttggggg 147
 Db 138 GTCTCGACTCTCTGGCTGGCTGATCAAGAGGAAGCAAGCAATGTCTGCTTTGGGA 197
 Qy 148 ctgcagctcacttgcaccatccgctgacagtcacagtgccgggtcagatttctctgg 207
 Db 198 CTGCAACTCCATACCTGATCATCTGCTGTAGTACAGTGGCCGAGTGTCTTGG 257
 Qy 208 gttccagccctctccagaagtgcagcggcgtggtcagctcttgaagacttagacaga 267
 Db 258 GGGCCAGCTTCTCCAGAGTGGCAGCAATGGCCAGCTACTAGGGACCTAGACAGGA 317
 Qy 268 gcacttcagaagttg-ttgaactttagtcggggccctgcatggaagactgcaga 326
 Db 318 GCACGCTTCAGAAAGTGTCTTGAAGTTTGTGTCAGCAGCCCTGCAGGGGAGATTGCCGA 377
 Qy 327 gaagctgtggagcaacttgggtccagcggcgaacctgtcagaagagcgtctggccctcctg 386
 Db 378 GAGTGTGCAGCGTNTTGGGGTTCAGGCCAACCTTCCGGAGGACAGTGTGCTT 437
 Qy 387 ctgg 390
 Db 438 GTTG 441

RESULT 3
 ID T32035
 AC T32035;
 DT 13-OCT-1996 (first entry)
 DE Proliferation-inhibiting gene sequence.
 KW Cell proliferation; cancer; viral infection; AIDS; HIV;

CC immunodeficiency virus, feline immunodeficiency virus, equine infectious
CC anemia virus, caprine arthritis encephalitis virus or visna virus.
CC Because the identified genes are non-essential to cell survival, the
CC treatment methods can be used in subjects without serious detrimental
CC effects to the subjects.
SQ Sequence 849 BP; 209 A; 169 C; 200 G; 216 T;

Query Match 3.6%; Score 39.8; DB 1; Length 849;
Best Local Similarity 57.1%; Pred. No. 0.07; 51; Indels 0; Gaps 0;
Matches 68; Conservative 0; Mismatches 51; Indels 0; Gaps 0;

Qy 938 ttgtttatagtcacacagctggcagggtggcacagtttataatctcagcccttgaag 997
Db 812 TGGGATTAGGAGGAAGTNNCTGGACATGGTTACATTTGGGTGTCATCTTAACACCTAGGAG 753
Qy 998 tctgagcgtggagaatgggaagtgaagctggcgtggcttctcagtgaggctcagtg 1056
Db 752 GCTGAGACAGGAGGATCACAAAGTTTAAAGGCTGACTGGCTGCATNTGATGCTCTGTG 694

RESULT 5
T88014 standard; DNA; 6727 BP.

ID T88014 standard; DNA; 6727 BP.
AC T88014:
DE Murine IL-5 cDNA genomic fusion gene.
KW Murine CD3delta; transcription control sequence; transgenic mouse; TCS;
KW T cell; interleukin-5; IL-5; thymocyte; disease model; therapeutic agent;
KW screening; eosinophil; chimeric; fusion gene; ss.
OS Chimeric - Mus sp.
PN WO9738086-A2.
PD 16-OCT-1997.
PF 09-APR-1997; U05932.
PR 09-APR-1996; US-629643.
PA (MAYO-) MAYO FOUNDATION.
PI Lee JJ, Lee NA;
DR WPI: 97-512706/47.
PT Transgenic mouse with interleukin-5 gene integrated into its genome
PT - under control of thymocyte and T cell, lung or basal keratinocyte
PT specific regulator, useful as disease model or to screen therapeutic
PT agent
PS Example 2; Page 73-75; 127pp; English.
CC This is a murine interleukin-5 (IL-5) cDNA genomic fusion gene. This
CC transgenic construct contain a chimeric pIL-5-EXPONA sequence, comprising
CC thymocyte/T cell, lung or basal keratinocyte specific transcription
CC control sequence (TCS) and a segment encoding interleukin-5 (IL-5), which
CC lacks endogenous 5'-control sequences when the TCS is thymocyte/T cell
CC specific. The TCS is usually from a murine CD3delta gene. The chimeric
CC DNA sequence is integrated into the genome and causes IL-5 to be
CC expressed at a level at least 50 times greater than the wild type. The mice can be
CC used as models of IL-5 associated disease, e.g. asthma, eosinophilic
CC pneumonia, myalgia, Loeffler's syndrome, allergy, emphysema, pulmonary
CC fibrosis, Wegener's granulomatosis, adult respiratory distress syndrome
CC (ARDS), bacterial or fungal infection, leukaemia, rheumatoid arthritis,
CC atopic dermatitis or contact hypersensitivity, or to screen for IL-5
CC antagonists which are potential therapeutic agents for an IL-5 associated
CC disease. These mice constitutively express IL-5 at high levels in
CC specific tissues.
SQ Sequence 6727 BP; 1765 A; 1538 C; 1414 G; 2010 T;

Query Match 3.4%; Score 37.4; DB 1; Length 6727;
Best Local Similarity 50.9%; Pred. No. 0.79;
Matches 89; Conservative 0; Mismatches 86; Indels 0; Gaps 0;

Qy 921 agcgggaacttcaggtttttttatagtcacagctggcagggtggcacagtttata 980
Db 1844 AGGTAGATATGAAAGTTTGTCTTATAATAAAGAAATATGAGATGGCAATATGTA 1903
Qy 981 atctcagcccttgaagctcagtgaggatgggaatgggaagtgaagctggcgtgcttc 1040

KW mole; wart; psoriasis; chromosome 4; gene therapy; diagnosis; ss.
OS Homo sapiens.
PN WO9622688-A1.
PD 01-AUG-1996.
PR 26-JAN-1996; U01344.
PA (BAYU) BAYLOR COLLEGE MEDICINE.
PI Bertram MJ, Ning Y, Pereira-Smith OM;
DR WPI: 96-362381/36.
PT Isolated human chromosome 4 fragment - used for therapy and
PT diagnosis of cell proliferative disorders, e.g. cancer, AIDS, moles,
PT warts and psoriasis.
PS Claim 4; Page 31a; 40pp; English.
CC 3 Human chromosome-4 fragments (T32033-35) comprise genetic
CC determinants that are capable of inhibiting or arresting the
CC proliferative capacity of a cell. They were identified using a
CC microcell-mediated chromosome transfer protocol. The fragments
CC are located at 4q26-28 or 4q33-34.1. They can be used in gene
CC therapy of cancer and diseases, such as AIDS, that involve
CC proliferative cells. They can also be used to diagnose the extent
CC and/or severity of malignancy.
SQ Sequence 218 BP; 69 A; 31 C; 63 G; 55 T;

Query Match 3.8%; Score 41.4; DB 1; Length 218;
Best Local Similarity 58.5%; Pred. No. 0.014;
Matches 72; Conservative 0; Mismatches 51; Indels 0; Gaps 0;

Qy 977 tataatctcagcccttgaagctcagcgtggagaatgggaagtgaagctggcctggc 1036
Db 36 TTTAATCCACCAATTGGGAGCAGGACGGCATATTTTGAGTTGAGGCCACCTGGT 95
Qy 1037 ttctcagtgaggctcagtgctgaattaaagcgaatgaagcaactattataaaaaa 1096
Db 96 TTACAGCTGAGTTTCAGACAGTCAGGGATATACAGAGAAACCTCTTCAGACAAAA 155
Qy 1097 aaa 1099
Db 156 AAA 158

RESULT 4
V06127/c standard; DNA; 849 BP.

ID V06127/c standard; DNA; 849 BP.
AC V06127:
DE Viral infection gene SEQ ID NO:47.
KW Viral infection; tumour suppressor; cellular gene; rat; cancer;
KW serum protein; inhibitor; malignant phenotype; HIV; influenza;
KW hepatitis; retrovirus; immunodeficiency; ds.
OS Rattus sp.
PN WO9739119-A1.
PD 23-OCT-1997.
PR 11-APR-1997; U06067.
PA (UYVA-) UNIV VANDERBILT.
PI Dubois RN, Organ EL, Rubin DH;
DR WPI: 97-526456/48.
PT Genes involved in viral infection and tumour suppression - used to
PT develop products for reducing or preventing viral infection or for
PT suppressing tumours
PS Claim 1; Page 64-65; 101pp; English.
CC The present sequence represents a viral infection gene. The present
CC invention describes nucleic acid sequences isolated from rat. The
CC sequences of the invention comprise 70 viral infection (VI) genes and
CC 8 tumour suppressor (TS) genes. Propagating cell cultures in the
CC absence of the serum protein (SP) allows selective elimination of cells
CC persistently infected with a virus from the cell culture. Inhibitors of
CC the SP can be used for reducing or inhibiting a viral infection.
CC Inhibitors of the TS gene products can be used to suppress a malignant
CC phenotype (MP). The methods and inhibitors can be used with viruses
CC such as HIV, influenza, hepatitis virus or animal retroviruses such as
CC simian immunodeficiency virus, avian immunodeficiency virus, bovine

```
Db 1904 ATCCAGTACTTGGGAGCCAAAGCGAGGCTAGTCTGAGTCTAGGCGCAGCTTAGATACA 1963
Qy 1041 atagtggcgtcagtgtaaatgaagaggtaagcaactattaaaaa 1095
Db 1964 TTGCCCTGTGTATCAAAAGTAAATCTCTATATAATAAACAACATTAGA 2018

RESULT 6
Q75270/c
ID Q75270 standard; DNA; 2858 BP.
AC Q75270;
DT 10-JUL-1995 (first entry)
DE Unique 2.9 kb genomic DNA fragment from human metastatic tumour.
KW metastasis; cancer; tumour cell; probe; diagnosis; identification;
OS malignant tumour; ds.
PN Homo sapiens.
PD W09428129-A.
PF 08-DEC-1994.
PR 27-MAY-1994; G01160.
PR 28-MAY-1993; GB-011130.
PA (ISIS-) ISIS INNOVATION LTD.
PI Tarin D;
DR WPI; 95-022801/03.
PT New human tumour metastasis gene - used to develop prods for
PT diagnosis or assessing the prognosis of tumour metastasis in a
PT patient.
PS Claim 1; Page 14-15; 22pp; English.
CC This sequence encodes a protein (sequence not given in the
CC specification) which is expressed in malignant human tumours and
CC their metastases. The DNA sequence has the potential to be a
CC valuable probe for the accurate assessment of the prognosis of
CC patients with malignant tumours, by examination of a tiny biopsy
CC sample, or even a few cells obtd. by fine needle aspiration, and
CC thus to influence therapy. No homology to any known coding regions
CC was found using EMBL and Genbank databases. Three possible exons
CC were predicted, one in the forward strand in frame 2 (between bases
CC 536 and 942), and two in the reverse strand, in frames 1 (between bases
CC 2143 and 2398) and 2 (between bases 1625 and 1907). PCR primers
CC were designed to be used with the RT-PCR technique to amplify homologous
CC mRNA sequences from RNA extracted from metastatic human tumour
CC cell lines (see Q75271 and Q79572-4).
SQ Sequence 2858 BP; 708 A; 763 C; 665 G; 722 T;

Query Match 3.3%; Score 36.6; DB 1; Length 2858;
Best Local Similarity 57.6%; Pred. No. 0.93;
Matches 87; Conservative 0; Mismatches 59; Indels 5; Gaps 1;

Qy 955 agctggcagggtggcagctttataatctcagcccttggaagttctgaggtggagaatg 1014
Db 2075 AGCCGGGGTGTGGCCCTACTGTGTTAATCCAGCACTCGGAAGCAGCAGGCGCAGATT 2016
Qy 1015 ggaagtgaagctggcctggctttcatagtgaggtcagtc-----gaattaaagag 1069
Db 2015 TCTGACTTATAGCCAGCTGTCTCAAAAGTGTGTTCCAGAACAGCCAGGCTACACAG 1956
Qy 1070 gtaagcaactattaaaaa 1100
Db 1955 AGAAACCTGTCTCAAAAAAACCAAAAAA 1925

RESULT 7
T65001/c
ID T65001 standard; DNA; 5687 BP.
AC T65001;
DT 28-MAY-1997 (first entry)
DE Mouse cell cycle regulatory protein SPA-1 genomic DNA (exons 1-4).
KW Cell division cycle; Span-N; Span-C; lymphocyte differentiation;
KW mouse; murine; ds.
OS Mus sp.
PN Key
DR Location/Qualifiers
FT exon 3109..3284
/*tag= a
```

```
FT exon /number= 1
FT 3764..4555
FT /*tag= b
FT exon /number= 2
FT 5147..5273
FT /*tag= c
FT exon /number= 3
FT 5383..5524
FT /*tag= d
FT /number= 4
PN J08217797-A.
PD 27-AUG-1996.
PF 17-FEB-1995; 051999.
PR 30-MAY-1994; JP-139513.
PR 20-OCT-1994; JP-279712.
PR 13-DEC-1994; JP-332520.
PA (MINA/) MINATO N.
DR WPI; 96-439574/44.
PT Gene encoding cell division-regulatory protein, SPA-1 - useful as
PT regulator of cell division and differentiation of lymphocytes
PS Claim 13; Page 12-14; 26pp; Japanese.
CC The present partial genomic DNA sequence comprises exons 1-4 for a
CC gene encoding cell division-regulating protein (designated SPA-1),
CC which is not expressed in the resting phase but is expressed after
CC the growth phase in the cell cycle of mammalian cells. SPA-1 is
CC useful as a regulator of cell division or for differentiation
CC of lymphocytes.
SQ Sequence 5687 BP; 1201 A; 1574 C; 1632 G; 1280 T;

Query Match 3.3%; Score 36.6; DB 1; Length 5687;
Best Local Similarity 56.1%; Pred. No. 1.2;
Matches 69; Conservative 0; Mismatches 54; Indels 0; Gaps 0;

Qy 977 tataatctcagccttggaagttctgaggtggagaatggaaagttgaagctggcctggc 1036
Db 2712 TTTAATCTAGCACTGGGAATCAGAGCGCAGCAGATTTCGAGTTCAAGCCAGCCTGT 2653
Qy 1037 ttcatagtgagctcagtgctgaataaagaggttaagcaactattaaaaa 1096
Db 2652 CTACAGAGTGTGATCCAGGATACCCAGGCTACACAGAGAAACCTGTCTTGAACAA 2593
Qy 1097 aaa 1099
Db 2592 AAA 2590

RESULT 8
X23315/c
ID X23315 standard; DNA; 4736 BP.
AC X23315;
DT 11-JUN-1999 (first entry)
DE Mouse 1-alpha-OHase promoter region DNA #1.
KW 1-alpha-OHase; vitamin D 1-alpha-hydroxylase; vitamin D pathway;
KW renal cytochrome P450 enzyme; 25-hydroxyvitamin D; human; mouse; rat;
KW diagnosis; treatment; disorder; endocrine; promoter; ss.
OS Mus sp.
PN W09907835-A2.
PD 18-FEB-1999.
PF 06-AUG-1998; CA0758.
PR 06-AUG-1997; US-906791.
PA (SHRI-) SHRINERS HOSPITALS FOR CHILDREN.
PI Glorieux FH, St-Arnaud R;
DR WPI; 99-190048/16.
PT New polynucleotides encoding 25-hydroxyvitamin D 1-alpha-hydroxylase -
PT useful to detect and treat vitamin D-related disorders
PS Example 4; Page 77-79; 85pp; English.
CC This invention describes novel 25-hydroxyvitamin D 1-alpha-hydroxylase
CC (1-alpha-OHase), a renal cytochrome P450 enzyme of the vitamin D
CC pathway, and polynucleotides encoding the enzyme from human, mouse and
CC rat tissue. Molecules of the invention are used to diagnose and treat
CC vitamin D-related disorders, to produce vitamin D metabolites, and to
CC identify modulators of the enzyme expression. The invention describes
```


[illegible]

FT		/label= Ap-1	/note= "transcription factor binding site"
FT	repeat_unit	2090. .2099	
FT		/*tag= v	
FT		/rpt_type= DIRECT	
FT		/label= I	
FT	repeat_unit	2133. .2142	
FT		/*tag= w	
FT		/rpt_type= DIRECT	
FT	protein_bind	/label= I	
FT		2144. .2152	
FT		/*tag= x	
FT		/label= Sp-1	
FT	protein_bind	/note= "transcription factor binding site"	
FT		2165. .2171	
FT		/*tag= y	
FT		/label= HNF-5	
FT	protein_bind	/note= "transcription factor binding site"	
FT		2325. .2332	
FT		/*tag= z	
FT		/label= Ap-2	
FT		/note= "transcription factor binding site"	
FT	protein_bind	2402. .2409	
FT		/*tag= aa	
FT		/label= Ap-2	
FT	protein_bind	/note= "transcription factor binding site"	
FT		2413. .2420	
FT		/*tag= ab	
FT		/label= Ap-1	
FT	repeat_region	/note= "transcription factor binding site"	
FT		2593. .2636	
FT		/*tag= ac	
FT	protein_bind	/rpt_type= DIRECT	
FT		2757. .2766	
FT		/*tag= ad	
FT		/label= p53	
FT	protein_bind	/note= "transcription factor binding site"	
FT		2776. .2784	
FT		/*tag= ae	
FT		/label= NF-IL6,Ap-1	
FT	protein_bind	/note= "transcription factor binding site"	
FT		2866. .2873	
FT		/*tag= af	
FT		/label= Ap-2	
FT	protein_bind	/note= "transcription factor binding site"	
FT		2908. .2919	
FT		/*tag= ag	
FT		/label= Myb,Mbf-1,Mep-1,Mtf-1	
FT	protein_bind	/note= "transcription factor binding site"	
FT		2926. .2933	
FT		/*tag= ah	
FT		/label= c/ebp	
FT	protein_bind	/note= "transcription factor binding site"	
FT		2951. .2956	
FT		/*tag= ai	
FT		/label= Pea-3	
FT	protein_bind	/note= "transcription factor binding site"	
FT		2988. .2993	
FT		/*tag= aj	
FT		/label= Myb	
FT	protein_bind	/note= "transcription factor binding site"	
FT		2998. .3003	
FT		/*tag= ak	
FT		/label= Pea-3	
FT	protein_bind	/note= "transcription factor binding site"	
FT		3076. .3086	
FT		/*tag= al	
FT		/label= SRE	
FT	protein_bind	/note= "transcription factor binding site"	
FT		3102. .3109	
FT		/*tag= am	
FT		/label= Ap-2	
FT		/note= "transcription factor binding site"	

FT	protein_bind	3162..3167	
FT		/tag= an	
FT		/label= Myb	
FT		/note= "transcription factor binding site"	
FT	protein_bind	3223..3228	
FT		/tag= go	
FT		/label= SRE	
FT		/note= "transcription factor binding site"	
FT	protein_bind	3256..3267	
FT		/tag= ap	
FT		/label= Sp-1,Ap-2,SRE	
FT		/note= "transcription factor binding site"	
FT	protein_bind	3306..3313	
FT		/tag= aq	
FT		/label= c/ebp	
FT		/note= "transcription factor binding site"	
FT	protein_bind	3352..3360	
FT		/tag= ar	
FT		/label= NF-IL6	
FT		/note= "transcription factor binding site"	
FT	repeat_unit	3466..3483	
FT		/tag= as	
FT		/rpt_type= DIRECT	
FT		/label= II	
FT	repeat_unit	3492..3509	
FT		/tag= at	
FT		/rpt_type= DIRECT	
FT	protein_bind	3507..3513	
FT		/label= III	
FT		/tag= au	
FT		/label= Ap-1	
FT		/note= "transcription factor binding site"	
FT	protein_bind	3702..3709	
FT		/tag= av	
FT		/label= Ap-2	
FT		/note= "transcription factor binding site"	
FT			
Query Match 3.3%; Score 36.2; DB 1; Length 51			
Best Local Similarity 60.8%; Pred. No. 1.5;			
Matches 59; Conservative 0; Mismatches 38; Indels			
QY	952	aacagctgggcagggtggcacagcttataatctcagcccttggaagctctgaggg	
Db	1071	AAAAATAGCGAGGTAGACAGTCCCTGTACCTCAGCGCTGAGGAAGTGAGGG	
QY	1012	atgggaagtgtaaagctgggccttgcttcatagtga 1048	
Db	1131	ATGGGAAGCTCAGACTGCTTCCTGGTTGCATGTTTAG 1167	
RESULT 13			
V05159/c			
ID	V05159	standard; DNA; 14180 BP.	
AC	V05159;		
DT	20-JUL-1998	(first entry)	
DE	Mouse butyrophilin gene.		
KW	Butyrophilin; btn gene; promoter; mouse; transgenic animal;		
KW	mammary gland-specific gene expression; breast cancer; diag		
KW	ds.		
OS	Mus musculus.		
PH	Key	Location/Qualifiers	
FT	promoter	1..4693	
FT		/tag= a	
FT		/note= "Claim 4"	
FT	TATA_signal	4645..4651	
FT		/tag= b	
FT	prim_transcript	4611	
FT		/tag= c	
FT		/evidence= EXPERIMENTAL	
FT		/standard_name= transcription start site	
FT	misc_signal	4645..4651	
FT		/note= "most frequently used site"	
FT		/tag= d	

```
FT FT misc_signal /note= "AT-rich, TATAA-like element"
FT FT /tag= e
FT FT /note= "AT-rich, TATAA-like element"
FT FT /tag= f
FT FT /note= "CCAAT-like element"
FT FT /tag= g
FT FT /note= "CCAAT-like element"
FT FT /tag= h
FT FT /note= "CCAAT-like element"
FT FT prim_transcript 4675
FT FT /tag= i
FT FT /evidence= EXPERIMENTAL
FT FT /standard_name= transcription start site
FT FT prim_transcript 4694
FT FT /tag= j
FT FT /evidence= EXPERIMENTAL
FT FT /standard_name= transcription start site
FT FT polyA_signal 13091..13096
FT FT /tag= k
FT FT /evidence= EXPERIMENTAL
FT FT prim_transcript 13097..13199
FT FT /tag= l
FT FT /evidence= EXPERIMENTAL
FT FT /standard_name= 3' end of transcript
FT FT CDS 4923..11397
FT FT /tag= m
FT FT /evidence= EXPERIMENTAL
FT FT /note= "contains introns"
FT FT sig_peptide 4923..5001
FT FT /tag= n
FT FT /evidence= EXPERIMENTAL
FT FT CDS 4650..4651
FT FT /tag= o
FT FT /evidence= EXPERIMENTAL
FT FT /standard_name= translational initiation codon
FT FT CDS 4743..4745
FT FT /tag= p
FT FT /evidence= EXPERIMENTAL
FT FT /standard_name= translational initiation codon
FT FT CDS 4765..4767
FT FT /tag= q
FT FT /evidence= EXPERIMENTAL
FT FT /standard_name= translational initiation codon
FT FT CDS 4776..4778
FT FT /tag= r
FT FT /evidence= EXPERIMENTAL
FT FT /standard_name= translational initiation codon
FT FT exon 4611..5001
FT FT /tag= s
FT FT /codon_start= 4923..4925
FT FT /number= 1
FT FT intron 5002..5520
FT FT /tag= t
FT FT /evidence= EXPERIMENTAL
FT FT /number= A
FT FT exon 5521..5871
FT FT /tag= u
FT FT /number= 2
FT FT intron 5872..8332
FT FT /tag= v
FT FT /evidence= EXPERIMENTAL
FT FT /number= B
FT FT exon 8333..8614
FT FT /tag= w
FT FT /number= 3
FT FT intron 8615..9485
FT FT /tag= x
FT FT /evidence= EXPERIMENTAL
FT FT /number= C
FT FT exon 9486..9635
FT FT /tag= y
FT FT /number= 4
FT FT /tag= z
FT FT /evidence= EXPERIMENTAL
FT FT /number= D
FT FT /tag= aa
FT FT /number= 5
FT FT /tag= ab
FT FT /evidence= EXPERIMENTAL
FT FT /number= E
FT FT /tag= ac
FT FT /number= 6
FT FT /tag= ad
FT FT /evidence= EXPERIMENTAL
FT FT /number= F
FT FT /tag= ae
FT FT /number= 7
FT FT repeat_unit 4807..4814
FT FT /tag= af
FT FT /rpt_type= INVERTED
FT FT repeat_unit 12556..12563
FT FT /tag= ag
FT FT /rpt_type= INVERTED
FT FT protein_bind 2948..2960
FT FT /tag= ah
FT FT /note= "NFI"
FT FT protein_bind complement (2971..2979)
FT FT /tag= ai
FT FT /note= "gamma-IRE"
FT FT protein_bind complement (3010..3015)
FT FT /tag= aj
FT FT /note= "myb"
FT FT protein_bind complement (3030..3036)
FT FT /tag= ak
FT FT /note= "YY1"
FT FT protein_bind complement (3115..3123)
FT FT /tag= al
FT FT /note= "gamma-IRE"
FT FT protein_bind 3129..3135
FT FT /tag= am
FT FT /note= "YY1"
FT FT protein_bind 3132..3146
FT FT /tag= an
FT FT /note= "NFI"
FT FT protein_bind 3138..3146
FT FT /tag= ao
FT FT /note= "STAT"
FT FT protein_bind complement (3145..3150)
FT FT /tag= ap
FT FT /note= "GMCSF"
FT FT protein_bind 3181..3189
FT FT /tag= aq
FT FT /note= "C/EBP"
FT FT protein_bind complement (3187..3197)
FT FT /tag= ar
FT FT /note= "C/EBP"
FT FT protein_bind 3190..3195
FT FT /tag= as
FT FT /note= "PU.1"
FT FT protein_bind complement (3194..3200)
FT FT /tag= at
FT FT /note= "alpha-IFN"
FT FT protein_bind 3195..3203
FT FT /tag= au
FT FT /note= "GMCSF"
FT FT protein_bind 3253..3259
FT FT /tag= av
```


PI Florence KA, Greene JM, Janat F, Lafleur DW, Ni J,
PI Rosen CA, Ruben SM, Shi Y, Young P, Yu G;
DR WPI: 99-264022/22.
DR P-PSDB: Y07861.
PT New isolated human genes and the secreted polypeptides they encode
PS Claim 1a: Page 232-233; 368pp; English.
CC This invention describes novel isolated human genes and the secreted
CC proteins they encode. The products of the invention are useful for
CC preventing, treating or ameliorating medical conditions, e.g. by protein
CC or gene therapy. Also pathological conditions can be diagnosed by
CC determining the amount of the new polypeptides in a sample or by
CC determining the presence of mutations in the new polynucleotides.
CC Specific uses are described for each of the 101 polynucleotides, based on
CC which tissues they are most highly expressed in, and include developing
CC products for the diagnosis or treatment of cancer, tumours,
CC neurodegenerative disorders, developmental abnormalities and fetal
CC deficiencies, blood disorders, leukemias, diseases of the immune system,
CC autoimmune diseases, hepatic and renal disease, lymphomas, inflammation,
CC allergies, Alzheimer's and cognitive disorders, schizophrenia, prostate
CC disease, skeletal or cardiac muscle disorders, pulmonary disorders,
CC transplant rejection, disorders involving osteoclasts such as
CC osteoporosis, arthritis or malignancies, digestive/endocrine disorders,
CC infections and AIDS. The human secreted proteins of the invention are
CC represented in Y07852-Y07993 and the encoding nucleic acids are
CC represented in X37451-X37552.
SQ Sequence 1089 BP; 367 A; 202 C; 240 G; 279 T;

Query Match 3.2%; Score 35.4; DB 1; Length 1089;
Best Local Similarity 69.6%; Pred. No. 1.4;
Matches 48; Conservative 0; Mismatches 21; Indels 0; Gaps 0;
QY 1030 gctggcttctagtgaggctcagtgctgaattaaagaggttaagcaactattaaaaa 1089
||||| || ||||| ||||| || ||||| ||||| ||||| ||||| ||||| |||||
Db 1015 GCCTGGGTGACAGGGTGAGACTCTGCTCAAAAAAAAAAAAAAAAAAAAAAAAAA 1074
QY 1090 aaaaaaaaaa 1098
|||||||
Db 1075 AAAAAAAAAA 1083

Search completed: May 15, 2000, 01:45:56
Job time: 9820 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 14, 2000, 22:48:01 ; Search time 1646.42 Seconds
(without alignments)
-649.938 Million cell updates/sec

Title: US-09-223-796-1
Perfect score: 1100
Sequence: 1 gcacgagccagccagcta.....attaaaaaaaaaaaaa 1100

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 882769 seqs, -486395729 residues

Total number of hits satisfying chosen parameters: 1765538

Minimum DB seq length: 0
Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database :

GenEmbl.*

- 1: gb_ba1.*
- 2: gb_ba2.*
- 3: gb_om.*
- 4: gb_ov.*
- 5: gb_pat.*
- 6: gb_ph.*
- 7: gb_pl1.*
- 8: gb_pl2.*
- 9: gb_pri.*
- 10: gb_pr2.*
- 11: gb_pr3.*
- 12: gb_ro.*
- 13: gb_sts.*
- 14: gb_sy.*
- 15: gb_un.*
- 16: gb_v1.*
- 17: em_fun.*
- 18: em_hum1.*
- 19: em_hum2.*
- 20: em_in.*
- 21: em_om.*
- 22: em_or.*
- 23: em_ov.*
- 24: em_pat.*
- 25: em_ph.*
- 26: em_pl.*
- 27: em_ro.*
- 28: em_sts.*
- 29: em_sy.*
- 30: em_un.*
- 31: em_v1.*
- 32: gb_htg1.*
- 33: gb_htg2.*
- 34: gb_in1.*
- 35: gb_in2.*
- 36: em_ba1.*
- 37: em_ba2.*
- 38: em_hum3.*
- 39: em_hum4.*
- 40: gb_pr4.*
- 41: gb_htg3.*
- 42: gb_htg4.*
- 43: gb_htg5.*
- 44: gb_htg6.*

- 45: gb_htg7.*
- 46: em_htg1.*
- 47: em_htg2.*
- 48: em_htg3.*
- 49: em_hum5.*
- 50: gb_pl3.*
- 51: gb_pr5.*
- 52: gb_htg8.*
- 53: gb_htg9.*
- 54: gb_htg10.*
- 55: gb_htg11.*
- 56: gb_htg12.*
- 57: gb_htg13.*
- 58: gb_htg14.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description
C 1	471.2	42.8	108369	40	AF124523	AF124523 Homo sapi
C 2	471.2	42.8	174923	52	AF216665	AF216665 Homo sapi
C 3	468	42.5	69528	40	AF146367	AF146367 Homo sapi
4	466.4	42.4	128009	40	AF118808	AF118808 Homo sapi
C 5	418.2	38.0	133200	43	AC011951	AC011951 Homo sapi
6	412.6	37.5	1588	40	AF113540	AF113540 Homo sapi
7	51.6	4.7	78543	44	AC011335	AC011335 Homo sapi
8	50.2	4.6	78651	12	AC007636	AC007636 Mus muscu
9	50.2	4.6	78783	12	AC007306	AC007306 Mus muscu
10	48.6	4.4	138608	56	AC022781	AC022781 Mus muscu
11	48.4	4.4	167859	55	AC016365	AC016365 Mus muscu
12	47.6	4.3	230475	42	AC015891	AC015891 Mus muscu
C 13	46.5	4.2	238554	55	AC020965	AC020965 Mus muscu
C 14	45.8	4.2	963	12	AF213391	AF213391 Mus muscu
C 15	45.6	4.1	7218	5	I66494	I66494 Sequence 14
C 16	45.6	4.1	247428	52	AC021632	AC021632 Mus muscu
C 17	45	4.1	139097	33	AC004591	AC004591 Mus muscu
18	45	4.1	171952	12	AC002324	AC002324 Mus muscu
19	44.8	4.1	82098	40	AC006252	AC006252 Homo sapi
C 20	44.8	4.1	207032	45	AC012224	AC012224 Homo sapi
C 21	44.6	4.1	172463	41	AC009948	AC009948 Homo sapi
22	44.6	4.1	182180	44	AC011404	AC011404 Homo sapi
23	44.6	4.1	191221	54	AC021629	AC021629 Mus muscu
C 24	44.6	4.1	233994	52	AC021627	AC021627 Mus muscu
C 25	44.4	4.0	934	12	MMU17267	U17267 Mus musculu
C 26	44.4	4.0	230475	42	AC015891	AC015891 Mus muscu
C 27	44.2	4.0	124479	55	AC020966	AC020966 Mus muscu
C 28	44.2	4.0	178964	43	AC012297	AC012297 Mus muscu
C 29	44.2	4.0	287292	55	AC020971	AC020971 Mus muscu
30	44	4.0	13390	12	AF030884	AF030884 Mus muscu
C 31	44	4.0	141662	55	AC006447	AC006447 Mus muscu
C 32	44	4.0	143586	55	AC006404	AC006404 Mus muscu
C 33	44	4.0	162038	53	AC019082	AC019082 Homo sapi
C 34	43.8	4.0	67473	55	AC023197	AC023197 Mus muscu
C 35	43.8	4.0	173740	55	AC020964	AC020964 Mus muscu
36	43.8	4.0	251364	12	MMAE000664	AE000664 Mus muscu
C 37	43.6	4.0	5006	12	MMU24171	U24171 Mus musculu
38	43.6	4.0	238554	55	AC020965	AC020965 Mus muscu
39	43.6	4.0	270294	55	AC020969	AC020969 Mus muscu
40	43.4	3.9	267	12	DNY16208	Y16208 Dryomys nit
41	43.4	3.9	189750	46	AC009860	AC009860 Homo sapi
42	43.4	3.9	206383	12	AC008160	AC008160 Mus muscu
C 43	43.2	3.9	269797	55	AC020968	AC020968 Mus muscu
44	43	3.9	37349	12	MUSXRCC1G	L34078 Mouse XRCC1
45	42.8	3.9	230383	55	AC021063	AC021063 Mus muscu

ALIGNMENTS

```
RESULT 1
AF124523/c AF124523 108369 bp DNA PRI 24-AUG-1999
LOCUS Homo sapiens Chromosome 8 clone Kox18-21 containing zinc-finger
DEFINITION protein 7 (ZFP7) gene, complete sequence.
ACCESSION AF124523
VERSION AF124523.1 GI:4325309
KEYWORDS HTG.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 108369)
AUTHORS Schudy,A., Rump,A., Drescher,B., Koczan,D., Thiesen,H.-J.,
Merck-Rousseau,M.F. and Rosenthal,A.
TITLE Direct Submission
JOURNAL Submitted (29-JAN-1999) Genome Analysis, Institute of Molecular
Biotechnology, Beutenbergstrasse 11, Jena 07745, Germany
FEATURES
source
1..108369
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="8"
1..245
/repeat_region
/rept_family="AluY"
/evidence-not_experimental
246..584
/repeat_region
/rept_family="L1MA4"
/evidence-not_experimental
585..863
/repeat_region
/rept_family="AluJo"
/evidence-not_experimental
864..1068
/repeat_region
/rept_family="L1MA4"
/evidence-not_experimental
1069..1353
/repeat_region
/rept_family="AluSx"
/evidence-not_experimental
1387..1734
/repeat_region
/rept_family="L1MA4"
/evidence-not_experimental
1735..2042
/repeat_region
/rept_family="AluY"
/evidence-not_experimental
2043..2172
/repeat_region
/rept_family="L1MA4"
/evidence-not_experimental
2173..2482
/repeat_region
/rept_family="AluSx"
/evidence-not_experimental
2483..2901
/repeat_region
/rept_family="L1MA4"
/evidence-not_experimental
2835..2912
/exon
/notes="GRAIL, score = 45.000%, comment = marginal"
2848..2883
/exon
/evidence-not_experimental
/notes="Xpound exon prediction, score = 77% (0%)"
2902..3202
/repeat_region
/rept_family="AluSp"
/evidence-not_experimental
3203..3325
/repeat_region
/rept_family="L1MA4"
/evidence-not_experimental
3326..3625
/repeat_region
/rept_family="AluSx"
/evidence-not_experimental
3626..3968
/repeat_region
/rept_family="L1MA4"
/evidence-not_experimental
4104..4431
/repeat_region
/rept_family="L2"
/evidence-not_experimental
4123..4369
/exon
/notes="GRAIL, score = 63.000%, comment = good"
4450..4608
/repeat_region
/rept_family="AluSg/x"
/evidence-not_experimental
4623..4796
/repeat_region
/rept_family="L2"
/evidence-not_experimental
4898..4997
/repeat_region
/rept_family="AluSg"
/evidence-not_experimental
5000..5060
/repeat_region
/rept_family="L1MA4A"
/evidence-not_experimental
5101..5402
/repeat_region
/rept_family="AluY"
/evidence-not_experimental
complement(5578..5595)
/repeat_region
/rept_family="MER33"
/evidence-not_experimental
5596..5867
/repeat_region
/rept_family="AluSx"
/evidence-not_experimental
5902..5912
/repeat_region
/rept_family="AluSx"
/evidence-not_experimental
complement(5913..6139)
/repeat_region
/rept_family="MER33"
/evidence-not_experimental
complement(8764..9075)
/repeat_region
/rept_family="AluSx"
/evidence-not_experimental
10935..11017
/repeat_region
/rept_family="MER74B"
/evidence-not_experimental
complement(11018..11306)
/repeat_region
/rept_family="AluJb"
/evidence-not_experimental
11307..11724
/repeat_region
/rept_family="MER74B"
/evidence-not_experimental
11725..12006
/repeat_region
/rept_family="AluSg"
/evidence-not_experimental
12049..12100
/repeat_region
/notes="homology = 100.00%, score = 26, counts = 2"
/rept_type=tandem
/rept_unit=actcaccttgagatgtgggtctgc
/evidence-not_experimental
complement(12081..12230)
/exon
/notes="GRAIL, score = 56.000%, comment = good"
/evidence-not_experimental
complement(12255..12748)
/repeat_region
/rept_family="LTR42"
/evidence-not_experimental
14255..14377
/exon
/notes="Genscan, score = 10.27%, comment = Initial_exon 123
bp frame: 1 phase: 0"
14386..14580
/exon
/evidence-not_experimental
/genes="zinc-finger protein 7 (ZFP7)"
number=1
join(14386..14580,15811..15958,16246..16372,23346..23462,
27285..29150)
/mRNA
/genes="zinc-finger protein 7 (ZFP7)"
/notes="mRNA acc.no. M29580"
/product="ZFP7"
14386..29150
/gene
/genes="zinc-finger protein 7 (ZFP7)"
```


* NOTE: This is a 'working draft' sequence.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.

FEATURES

source	Location/Qualifiers	
	1. .174923	
	/organism="Homo sapiens"	
	/db_xref="taxon:9606"	
	/chromosome="8"	
	/map="8q24"	
	/clone="BAC 704B241056"	
	/complement(36..349)	
	/rpt_family="AluSx"	
	352..405	
	/rpt_type=tandem	
	/complement(449..582)	
	/note="FEXHB"	
	/evidence=not_experimental	
	/complement(457..748)	
	/rpt_family="AluSg"	
	/complement(750..820)	
	/rpt_family="t1M4_orf2"	
	/complement(1167..1754)	
	/rpt_family="t1M4_orf2"	
	1387..1410	
	/rpt_type=tandem	
	/complement(1410..1713)	
	/rpt_family="AluY"	
	1434..1555	
	/note="FEXHB"	
	/evidence=not_experimental	
	1517..1702	
	/rpt_type=inverted	
	/complement(1865..2170)	
	/rpt_family="AluJo"	
	2386..2461	
	/rpt_type=tandem	
	/complement(2435..2544)	
	/rpt_family="AluJb"	
	/complement(2545..2649)	
	/rpt_family="AluSx"	
	2738..3017	
	/rpt_family="AluJo"	
	2747..2931	
	/rpt_type=inverted	
	3369..3458	
	/note="GENSCAN"	
	4066..4182	
	/note="MZEf"	
	/evidence=not_experimental	
	/complement(4603..5038)	
	/rpt_family="MLTIC"	
	/complement(4836..4889)	
	/note="GRAIL"	
	/evidence=not_experimental	
	/complement(5199..5235)	
	/note="GRAIL"	
	/evidence=not_experimental	
	/complement(5199..5235)	
	/note="MZEf"	
	/evidence=not_experimental	
	5336..5612	
	/rpt_family="AluY"	
	5472..5530	
	/rpt_type=inverted	
	5627..5906	
	/rpt_type=tandem	
	/complement(6589..6751)	
	/rpt_family="AluSg"	
	6689..6747	
	/rpt_type=inverted	

exon	8005..8268	/note="MZEf"
	/evidence=not_experimental	
exon	8005..9818	
	/note="GENSCAN"	
exon	8046..9818	
	/note="FEXHB"	
	/evidence=not_experimental	
exon	8046..9818	
	/note="GRAIL"	
	/evidence=not_experimental	
exon	/complement(9207..9412)	
	/note="GRAIL"	
	/evidence=not_experimental	
repeat_region	9299..9578	
	/rpt_type=tandem	
repeat_region	9958..10259	
	/rpt_family="AluY"	
exon	/complement(10113..10232)	
	/note="FEXHB"	
	/evidence=not_experimental	
repeat_region	10238..10259	
	/rpt_type=tandem	
repeat_region	11078..11384	
	/rpt_family="AluY"	
repeat_region	11359..11384	
	/rpt_type=tandem	
repeat_region	11519..11638	
	/rpt_type=tandem	
repeat_region	/complement(11792..11848)	
	/rpt_family="LTRA0b"	
exon	/complement(11834..11860)	
	/note="GENSCAN"	
repeat_region	/complement(12208..12332)	
	/rpt_family="HERVL40"	
exon	/complement(12368..12424)	
	/note="GENSCAN"	
exon	/complement(12368..12424)	
	/note="XPOUND"	
	/evidence=not_experimental	
exon	/complement(12368..12468)	
	/note="GRAIL"	
	/evidence=not_experimental	
exon	/complement(12673..12773)	
	/note="GRAIL"	
	/evidence=not_experimental	
repeat_region	12872..12961	
	/rpt_family="HERVL40"	
repeat_region	12985..13069	
	/rpt_family="MERS58B"	
repeat_region	13354..13491	
	/rpt_type=inverted	
repeat_region	/complement(13354..13641)	
	/rpt_family="AluY"	
repeat_region	13949..14111	
	/rpt_family="FRAM"	
repeat_region	14263..14357	
	/rpt_family="L2b"	
exon	/complement(14534..14650)	
	/note="GRAIL"	
	/evidence=not_experimental	
repeat_region	14715..14969	
	/rpt_family="AluJo"	
repeat_region	15002..15162	
	/rpt_family="AluSg"	
repeat_region	15022..15161	
	/rpt_type=inverted	
repeat_region	15168..15612	
	/rpt_family="MERS58A"	
repeat_region	16383..16634	
	/rpt_family="L2"	
exon	16705..16831	
	/note="GRAIL"	

[illegible]


```
* 5793 6853: contig of 1061 bp in length
*      gap of unknown length
* 6854 8428: contig of 1575 bp in length
*      gap of unknown length
* 8429 9591: contig of 1163 bp in length
*      gap of unknown length
* 9592 10969: contig of 1378 bp in length
*      gap of unknown length
* 10970 12475: contig of 1506 bp in length
*      gap of unknown length
* 12476 14088: contig of 1613 bp in length
*      gap of unknown length
* 14089 15243: contig of 1155 bp in length
*      gap of unknown length
* 15244 16417: contig of 1174 bp in length
*      gap of unknown length
* 16418 17553: contig of 1136 bp in length
*      gap of unknown length
* 17554 18961: contig of 1408 bp in length
*      gap of unknown length
* 18962 20529: contig of 1568 bp in length
*      gap of unknown length
* 20530 23589: contig of 3060 bp in length
*      gap of unknown length
* 23590 25249: contig of 1660 bp in length
*      gap of unknown length
* 25250 26988: contig of 1739 bp in length
*      gap of unknown length
* 26989 28426: contig of 1438 bp in length
*      gap of unknown length
* 28427 29981: contig of 1555 bp in length
*      gap of unknown length
* 29982 31552: contig of 1571 bp in length
*      gap of unknown length
* 31553 33219: contig of 1667 bp in length
*      gap of unknown length
* 33220 35694: contig of 2475 bp in length
*      gap of unknown length
* 35695 37968: contig of 2274 bp in length
*      gap of unknown length
* 37969 39872: contig of 1904 bp in length
*      gap of unknown length
* 39873 42198: contig of 2326 bp in length
*      gap of unknown length
* 42199 43732: contig of 1534 bp in length
*      gap of unknown length
* 43733 45927: contig of 2195 bp in length
*      gap of unknown length
* 45928 47914: contig of 1987 bp in length
*      gap of unknown length
* 47915 50000: contig of 2086 bp in length
*      gap of unknown length
* 50001 51834: contig of 1834 bp in length
*      gap of unknown length
* 51835 54069: contig of 2235 bp in length
*      gap of unknown length
* 54070 56363: contig of 2294 bp in length
*      gap of unknown length
* 56364 59354: contig of 2991 bp in length
*      gap of unknown length
* 59355 63071: contig of 3717 bp in length
*      gap of unknown length
* 63072 65969: contig of 2898 bp in length
*      gap of unknown length
* 65970 69646: contig of 3677 bp in length
*      gap of unknown length
* 69647 73243: contig of 3597 bp in length
*      gap of unknown length
* 73244 76844: contig of 3601 bp in length
*      gap of unknown length
* 76845 79797: contig of 2953 bp in length
*      gap of unknown length
* 79798 83127: contig of 3330 bp in length
```

```
* 83128 85980: gap of unknown length
*      contig of 2855 bp in length
* 85983 89291: gap of unknown length
*      contig of 3309 bp in length
* 89292 92357: gap of unknown length
*      contig of 3066 bp in length
* 92358 95772: gap of unknown length
*      contig of 3415 bp in length
* 95773 99411: gap of unknown length
*      contig of 3639 bp in length
* 99412 103446: gap of unknown length
*      contig of 4035 bp in length
* 103447 108787: gap of unknown length
*      contig of 5341 bp in length
* 108788 116339: gap of unknown length
*      contig of 7552 bp in length
* 116340 124476: gap of unknown length
*      contig of 8137 bp in length
* 124477 133200: gap of unknown length
*      contig of 8724 bp in length.
FEATURES             Location/Qualifiers
     source            1..133200
                        /organism="Homo sapiens"
                        /db_xref="taxon:9606"
                        /clone_lib="RP11-18M17"
                        /clone_lib="RPC1-11 Human Male BAC"
BASE COUNT  38324 a 27723 c 27998 g 39103 t 52 others
ORIGIN
```

```
Query Match      38.0%; Score 418.2; DB 43; Length 133200;
Best Local Similarity 77.2%; Pred. No. 1.9e-110;
Matches 521; Conservative 0; Mismatches 153; Indels 1; Gaps 1;
```

```
Qy 133 tgctgtttgggggtgcagctccatactgcaacctcccgctgaagtcacagtcacagtcgccc 192
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 99292 TgTTTGTGTGGGGGCCACAGCTCTGTA-TTGCATCATCTCTGGTGTGTAGTCACAGTGGCC 99234

Qy 193 gggtcagtttctctgggtcccgccctctccagaagtgcagccgtggtcagctctcttga 252
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 99233 GAGTGAAGTTTTGGGGGCCAGCTTCTCCAGAGGTGGCAGCAATGCACGACTACTAG 99174

Qy 253 aggactagacagagacaccttcagaagtgttgaaactgtagtctggggccctgcagt 312
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 99173 GGGACTTAAACAGGAGCATGTTTAGAAGTTGCTGAAGTTTGTGGTCAGCAGCTGCAAG 99114

Qy 313 ggaagactgcagagaagctgtgagcaacttggccagcgccaaacctgtcagaagagc 372
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 99113 GGGAGGAGTCCCGAGAGGCTGTGCAGTGTCTTGGGGTCAGCACCACTGCCGAGGAGT 99054

Qy 373 gtcgtggcgtctctgtgcggggcacacacacccctgtctccagcaggtctctcggctgccc 432
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 99053 GGCCTGGTGCCCTACTGCGAGGCATGCACAGCTGCTCCAGCAGGCCCTCCATCTTCCCC 98994

Qy 433 ctgtagtctaaagccagatgccttcagagaagagctccaggaacttggcaattcctcagg 492
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 98993 CCACACGCTGAAGCCCAACACCTTCAGGGACCAGCTCCAGGAACCTGTGATCCCCCAAG 98934

Qy 493 atctaattgagatttggccagtttggcatttggagtaacacctctctcagactctg 552
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 98933 ACCTGGTCGGGGACTTGGCCAGCGCTGTATTTGGGAGCCAGCGGCCCTCTTGTATTCGT 98874

Qy 553 tagcccaacagcaggatctctcgtcctcacgtgtcttactccggtggtggtggtgacg 612
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 98873 TGGCCCCAGCAGCAGGGGCGCTGGGTGCCCATGTTGTGACTTTTCAGTGGTGGGTGATG 98814

Qy 613 tggccatctcaaccagcgtcagtcctcctcctcctcagcaaccagagtggttctcatgcagtga 672
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 98813 TGACAATCTCCACACAGTGCCCTCGCTTCGCTTCGTCGACCGCGGGTCTGTATGACAGTGA 98754

Qy 673 agctcacagatggtctgcacacccgctcaggtgcccataagccaaatttcaggagctgc 732
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 98753 AGCTTTTCAGATGGGTAGCATACCGCTTCGAGGGTCCCCCATAGCCCAAGTTCAGGAGGTGC 98694
```

```

QY 733 ggtcagtgtagccttgcttaagagagatgcagaaactggagaagaagtgtgagcgca 792
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 98693 GGTACAGCGTGGCCCTGGTCTCTAAAGAGATGCGAGATCTGGAGAGAGGTGTGAGGGCA 98634

QY 793 aactgcagagactgac 807
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 98633 GCCTGCAGGACTGAC 98619

RESULT 6
AF113540 1588 bp mRNA PRI 04-DEC-1999
LOCUS Homo sapiens hypothalamus protein HT002 mRNA, complete cds.
DEFINITION AF113540
VERSION AF113540.1 GI:6523834
KEYWORDS human.
SOURCE Homo sapiens
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 1588)
AUTHORS Dong, H., Ren, S., Huang, C., Jiang, C., Li, Y., Zhou, J., Yu, Y., Xu, S.,
Wang, Y., Fu, G., Chen, Z. and Han, Z.
TITLE A novel gene expressed in the human hypothalamus
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 1588)
AUTHORS Dong, H., Ren, S., Huang, C., Jiang, C., Li, Y., Zhou, J., Yu, Y., Xu, S.,
Wang, Y., Fu, G., Chen, Z. and Han, Z.
DIRECT SUBMISSION
SUBMITTED (16-DEC-1998) Chinese National Human Genome Center at
Shanghai, 351 Guo Shoujing Rd., Zhangjiang Hi-Tech Park, Pudong,
Shanghai 201203, China
FEATURES
Source Location/Qualifiers
1..1588
/organism="Homo sapiens"
/db_xref="taxon:9606"
/tissue_type="hypothalamus"
418..1089
/codon_start=1
/product="hypothalamus protein HT002"
/protein_id="AAFL4877.1"
/db_xref="GI:6523835"
/translation="MSAVGTATPYLHPGDSHGRVSLGALQLPPEVAMARLLGLDLD
RSTFKLLFVYSLQGEDCRDCSVLASAPTCRAAGCLLAGMTLLQOALRLPPTS
LKPDTPDPAPCALHPPRGLGIGOWEPAALLDSVAQOQAMLPVADPRWRVDV
ATISALARSLOPSVLMLQLSDGSAYREVFTAKFQLRYSLVALVKEMADLEKRC
RRLQD"
BASE COUNT 354 a 437 c 437 g 360 t
ORIGIN

Query Match 37.5%; Score 412.6; DB 40; Length 1588;
Best Local Similarity 75.8%; Pred. No. 5.6e-109;
Matches 562; Conservative 0; Mismatches 174; Indels 5; Gaps 4;

QY 84 tgtgttagagacctctcttcgagctgggtcagaagaacggagaagaggtgtctgttg 143
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 370 TCTGGGACCGCACTCTCTGGCTGGCTGATCAAGAGAGAGACCAACATGCTGCTGTG 429

QY 144 ggggtgcagctccactgtaccatcccgcgtgacagtcacagtgcccggtcagtttc 203
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 430 GGGACTGCAACTCATACCTGTCATCTCTGGTGATGATGATGATGATGATGATGATGAT 489

QY 204 ctgggttccagccctctccagaagtgacggcggtggtcagctcttgaagagactagac 263
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 490 TTGGGGCCGAGCTTCTCCAGAGGTGGCAGCAATGCCCGGCTACTAGGGACCTAGAC 549

QY 264 agagacaccttcagaagttgtgaactgttagtgcggccctcatggtggaagactgc 323
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 550 AGGAGCAGCTTCAGAAAGTGTGAAGTTGGTGTGAGCAGCTGAGGGGAGGAGACTGC 609

QY 324 agagaagctgtgagcaacttggtgcccagcgcccaacctgtcagaagaagcgtctgcccgc 383

```

```

Db 610 CGAGACTGT-TGCAGCGTCTTGGGGTCAGCGCCAACTCGCGGAGAGCAGCTGGG--TGC 666
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 384 ctgctgcgggcacacacacacacctgtctccagcaggctctccggctgcccctgctagtcta 443
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 667 CTGCTGGCAGGATGCACACTGCTCCAGCAGGCGCTCCGTCTGCCCGCCACACGCGCTG 736
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 444 aagccagatgctctccaggga-agagctccaggaaacttgcattccctcaggatctaattgg 502
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 727 AAGCCTGACACCTTCAGGGACCCAGCTCCAGGAGCTCTGCATCCCCCAAGACCTGTGTCGG 786
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 503 agattggccagtttgatctggagtgagtgagtgagtgagtgagtgagtgagtgagtgag 562
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 787 GGACTTGGCCAGCGTGGTATTTGGGAGCCAGCGGCC-CTCCTTGATTCTGTGGCCCGCAGCA 845
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 563 gaaggatcctgcctcctcagctgtcttaactccggtggtggcggtggaagctggaatctc 622
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 846 GCAGGGGGCTGGCTGCCGATGTTGCTGACTTTCGGTGGCGGTGGATGTAGCAATCTC 905
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 623 aaccagctcagtcctccctgctccctgcaaccgagtggtctctcagctcagctcagctcaga 682
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 906 CACCAGTGGCTGGCTCGCTCCCTGCGAGCGGAGCGTCTGTGATGCAGCTGAAGCTTTCAGA 955
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 683 tggatctcacaccgcttcgagtggtgcccataagcgaatttcaggagctgcggtacagtg 742
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 966 TGGGTGAGCATACCGCTTTGAGGTCCCGACAGCCAAAGTTCCAGGAGCTGCGGTACAGCGT 1025
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 743 agccttggtccttaagagatgagcagaactgagagaagaagtgtgagcgaactcagcaga 802
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1026 GGCCTTGGTCTTAAAGAGAGATGCGAGATCTGGAGAAAGAGTGTGAGCGGAGAGTGCAGGA 1085
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 803 ctgactgaaccctggactgt 823
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1086 CTGACCCCTCACTTGACCACT 1106
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

RESULT 7

```

LOCUS AC011335 78543 bp DNA HTG 19-DEC-1999
DEFINITION Homo sapiens chromosome 5 clone CIT-HSPC_235N22, WORKING DRAFT
SEQUENCE, 27 unordered pieces.
ACCESSION AC011335
VERSION AC011335.2 GI:6604456
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 78543)
AUTHORS DOE Joint Genome Institute.
TITLE Sequencing of Human Chromosome 5
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 78543)
AUTHORS DOE Joint Genome Institute.
TITLE Direct Submission
JOURNAL Submitted (06-OCT-1999) Production Sequencing Facility, DOE Joint
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
COMMENT On Dec 20, 1999 this sequence version replaced gi:6013605.
-----Genome Center
Center: Joint Genome Institute
Web site: http://www.jgi.doe.gov
-----Summary Statistics
Estimated insert size: 78543; sum-of-contigs estimation
Estimated insert size: 38000; pulse field gel estimation
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 27 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

```

Db	39350	TCTGAGTTCAGCCGAGCCTGGGCTATACAGAGAAACCCCTGTCTCAAAAAACCAAAAAA	39409
Qy	1075	gcaactattaaaaaaaaaaaaaaaaa 1100 	
Db	39410	AAAAAAAAAAAAAAAAAGRAAAAGAA 39435	
RESULT	8		
LOCUS	AC007636	78651 bp DNA	15-FEB-2000
DEFINITION	Mus musculus chromosome 6 clone unknown strain RIII Fibroblast cell line C127, complete sequence.		
ACCESSION	AC007636		
VERSION	AC007636.19	GI:6862971	
KEYWORDS	HTG.		
SOURCE	house mouse.		
ORGANISM	Mus musculus		
REFERENCE	Eukaryota; Metazoa; Chordata; Creniata; Vertebrata; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.		
AUTHORS	1 (bases 1 to 78651)		
TITLE	Yao,Z., Shual,S., Wu,H., Meisler,M. and Roe,B.A.		
JOURNAL	Mus musculus Chromosome 6 PAC Clone pl_4 In MND2 Region Unpublished		
REFERENCE	2 (bases 1 to 78651)		
AUTHORS	Yao,Z., Shual,S., Wu,H., Meisler,M. and Roe,B.A.		
TITLE	Direct Submission		
JOURNAL	Submitted (21-MAY-1999) Department of Chemistry And Biochemistry, The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman, OK 73019, USA		
REFERENCE	3 (bases 1 to 78651)		
AUTHORS	Yao,Z., Shual,S., Wu,H., Meisler,M. and Roe,B.A.		
TITLE	Direct Submission		
JOURNAL	Submitted (03-FEB-2000) Department of Chemistry And Biochemistry, The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman, OK 73019, USA		
REFERENCE	4 (bases 1 to 78651)		
AUTHORS	Yao,Z., Shual,S., Wu,H., Meisler,M. and Roe,B.A.		
TITLE	Direct Submission		
JOURNAL	Submitted (15-FEB-2000) Department of Chemistry And Biochemistry, The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman, OK 73019, USA		
COMMENT	On Feb 3, 2000 this sequence version replaced gi:6730773.		
FEATURES	Location/Qualifiers		
source	1..78651		
	/organism="Mus musculus"		
	/strain="RIII Fibroblast cell line C127"		
	/db_xref="taxon:10090"		
	/chromosome="6"		
	/clone="unknown"		
	/clone_1111="Nat Sternberg mouse pl RIII"		
BASE COUNT	20921 a 20258 c 18946 g 18526 t		
ORIGIN			
Query Match	4.6%;	Score 50.2;	DB 12; Length 78651;
Best Local Similarity	59.4%;	Pred. No. 0.0015;	
Matches	85;	Conservative 0;	Mismatches 58; Indels 0; Gaps 0;
Qy	955	agctggcgaggtggcacagtttatactcagcccttggagctcaggtcaggtcgagaatg 1014 	
Db	52591	AAC7GGCTCTGGTGGAGGGCTAGAACCCCGAGCCACTGACGACTGACAGAGGAGATC 52650 	
Qy	1015	ggaaagtgaagctggcgctcttcagttaggctcaggtcgaattaaaggaggtaaa 1074 	
Db	52651	ACAAAGTTCAGAGCCACCTGGGCGAACTACTGAGACCTGTCTCGAAAAACAAAC 52710 	
Qy	1075	gcaactattaaaaaaaaaaaaaaaaa 1097 	
Db	52711	AAACATTAATAATAATAATAA 52733	
RESULT	9		
AC007306			

LOCUS AC007306 78783 bp DNA ROD 15-FEB-2000
DEFINITION Mus musculus chromosome 6 clone unknown strain RIII fibroblast cell
line C127, complete sequence.
ACCESSION AC007306
VERSION AC007306.15 GI:6958029
KEYWORDS HTG
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE
AUTHORS Chen, F., Do, T., Do, A., Meisler, M. and Roe, B. A.
TITLE Mus musculus Chromosome 6 PF Clone pl_2 In MND2 Region
JOURNAL Unpublished
AUTHORS
REFERENCE 2 (bases 1 to 78783)
AUTHORS Chen, F., Do, T., Do, A., Meisler, M. and Roe, B. A.
TITLE Direct Submission
JOURNAL Submitted (14-APR-1999) Department of Chemistry And Biochemistry,
The University of Oklahoma, 620 Farrington Oval, Room 208, Norman,
OK 73019, USA
REFERENCE 3 (bases 1 to 78783)
AUTHORS Chen, F., Do, T., Do, A., Meisler, M. and Roe, B. A.
TITLE Direct Submission
JOURNAL Submitted (10-FEB-2000) Department of Chemistry And Biochemistry,
The University of Oklahoma, 620 Farrington Oval, Room 208, Norman,
OK 73019, USA
REFERENCE 4 (bases 1 to 78783)
AUTHORS Chen, F., Do, T., Do, A., Meisler, M. and Roe, B. A.
TITLE Direct Submission
JOURNAL Submitted (15-FEB-2000) Department of Chemistry And Biochemistry,
The University of Oklahoma, 620 Farrington Oval, Room 208, Norman,
OK 73019, USA
COMMENT On Feb 11, 2000 this sequence version replaced gi:6691340.
FEATURES
source location/Qualifiers
1..78783
/organism="Mus musculus"
/strain="RIII Fibroblast cell line C127"
/db_xref="taxon:10090"
/chromosome="6"
/clone="unknown"
/clone.lib="Nat Sternberg mouse P1 RIII"
BASE COUNT 20285 a 21224 c 19078 g 18156 t
ORIGIN
Query Match 4.6%; Score 50.2; DB 12; Length 78783;
Best Local Similarity 59.4%; Pred. No. 0.0015;
Matches 85; Conservative 0; Mismatches 58; Indels 0; Gaps 0;
QY 955 agctggcagggtggcacagtttataatcagcccttggaagctcgtgagctggagaatg 1014
Db 20099 AACTGGCTCTGTGTGAGAGGCGCTAGAACCCGACCTGAGAGCTGAGAGGAGAAATC 20158
QY 1015 ggaagtgaagctggcctgtttcctagtgagcctcagtgctgaattaaagaggttaa 1074
Db 20159 ACAAGTTCAAGCCCACTGGCGCACTTACTGAGACCTTGCTCGAAACAAACAAAC 20218
QY 1075 gcaactattaaaaa 1097
Db 20219 AAAACATTAAAAATAATATAAA 20241
RESULT 10
AC022781
LOCUS AC022781 138608 bp DNA HTG 18-FEB-2000
DEFINITION Mus musculus chromosome 11 clone RP23-199H17 map 11, LOW-PASS
SEQUENCE SAMPLING.
ACCESSION AC022781
VERSION AC022781.2 GI:6997225
KEYWORDS HTG; HTGS_PHASE0.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE
AUTHORS Birren, B., Linton, L., Nusbaum, C., Lander, E., Abraham, H., Allen, N.,
Boguslavskiy, L., Bouckhalter, B., Brown, A., Burkett, G., Castle, A.,
Choepel, I., Colangelo, M., Collins, S., Collymore, A., Cooke, P.,
Dearellano, K., Dewar, K., Domino, M., Doyle, M., Fenestor, J.,
Ferrelira, P., FitzHugh, W., Forrest, C., Gage, D., Galagan, J.,
Gardyna, S., Grant, G., Hagos, B., Heaford, A., Horton, L.,
Howard, J. C., Johnson, R., Jones, C., Kann, L., Karatas, A., Klein, J.,
Landers, T., Lehoczy, J., Levine, R., Liu, C., Liu, G., Locke, K.,
Macdonald, P., Marquis, N., McEwan, P., McQuirk, A., McKernan, K.,
McPheders, R., Meldrum, J., Meneus, L., Morrow, J., Naylor, J.,
Norman, C. H., O'Connor, T., O'Donnell, P., Olivari, T. M., Peterson, K.,
Pierre, N., Pisan, C., Pollara, V., Raymond, C., Riley, R., Rothman, D.,
Roy, A., Santos, R., Severy, P., Spencer, B., Stange-Thomann, N.,
Stojanovic, N., Subramanian, A., Talamas, J., Tesfaye, S., Theodore, J.,
Tirrell, A., Vassiliev, H., Viel, R., Vo, A., Wu, X., Wyman, D., Ye, W. J.,
Zimmer, A. and Zody, J. M.
TITLE Direct Submission
JOURNAL Submitted (06-FEB-2000) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT On Feb 18, 2000 this sequence version replaced gi:6922475.
All repeats were identified using RepeatMasker:
Smit, A. F. A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L5913
Center clone name: 199_H_17

* NOTE: This record contains 183 individual
* sequencing reads that have not been assembled into
* contigs. Runs of N are used to separate the reads
* and the order in which they appear is completely
* arbitrary. Low-pass sequence sampling is useful for
* identifying clones that may be generic and allows
* overlap relationships among clones to be deduced.
* However, it should not be assumed that this clone
* will be sequenced to completion. In the event that
* the record is updated, the accession number will
* be preserved.
1 798: contig of 798 bp in length
1 799: gap of unknown length
799 1617: contig of 819 bp in length
1618 2414: contig of 797 bp in length
2415 3240: gap of unknown length
3241 4035: gap of unknown length
4036 4857: contig of 822 bp in length
4858 5686: gap of unknown length
5687 6490: contig of 804 bp in length
6491 7300: gap of unknown length
7301 8124: contig of 810 bp in length
8125 8943: gap of unknown length
8944 9746: contig of 803 bp in length

ACCESSION	AC015891	SEQUENCE SAMPLING.			
VERSION	AC015891.1	GI-6446777			
KEYWORDS	HIG; HTGS_PHASE0; NULL.				
SOURCE	house mouse.				
ORGANISM	Mus musculus				
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;				
AUTHORS	Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.				
TITLE	1 (bases 1 to 230475)				
REFERENCE	Birren, B., Linton, L., Nusbaum, C., Lander, E., Allen, N., Anderson, M.,				
AUTHORS	Birren, B., Linton, L., Nusbaum, C., Lander, E., Allen, N., Anderson, M.,				
JOURNAL	Baldwin, J., Barna, N., Beckerly, R., Boguslavsky, L., Boukhgalter, B.,				
REFERENCE	Brown, A., Castle, A., Colangelo, M., Collins, S., Collymore, A.,				
AUTHORS	Cooke, P., Dearellano, K., Dewar, K., Domino, M., Donelan, L., Doyle, M.,				
	Ferreira, P., Fitzhugh, W., Forrest, C., Funke, R., Gage, D.,				
	Galagan, J., Gardyna, S., Grant, G., Hegos, B., Hearford, A., Horton, L.,				
	Howland, J. C., Johnson, R., Jones, C., Kann, L., Karatas, A., Klein, J.,				
	Lehoczky, J., Lieu, C., Locke, K., Macdonald, P., Marquis, N.,				
	McEwan, P., McGuck, A., McKernan, K., McLaughlin, J., Meldrim, J.,				
	Morrow, J., Naylor, J., Norman, C. H., O'Connor, T., O'Donnell, P.,				
	Peterson, K., Pollara, V., Riley, R., Roy, A., Santos, R., Severy, P.,				
	Stange-Thomann, N., Stojanovic, N., Subramanian, A., Talamas, J.,				
	Tesfaye, S., Tirrell, A., Vassiliev, H., Vo, A., Wheeler, J., Wu, X.,				
	Wyman, D., Ye, W. J., Zimmer, A. and Zody, M.				
TITLE	Direct Submission				
JOURNAL	Submitted (17-NOV-1999) Whitehead Institute/MIT Center for Genome				
COMMENT	Research, 320 Charles Street, Cambridge, MA 02141, USA				
	All repeats were identified using RepeatMasker:				
	Smit, A.F.A. & Green, P. (1996-1997)				
	http://ftp.genome.washington.edu/RM/RepeatMasker.html				
	----- Project Information				
	Center project name: l761				
	Center clone name: 522_L_13				

	* NOTE: This record contains 279 individual				
	* sequencing reads that have not been assembled into				
	* contigs. Runs of N are used to separate the reads				
	* and the order in which they appear is completely				
	* arbitrary. Low-pass sequence sampling is useful for				
	* identifying clones that may be gene-rich and allows				
	* overlap relationships among clones to be deduced.				
	* However, it should not be assumed that this clone				
	* will be sequenced to completion. In the event that				
	* the record is updated, the accession number will				
	* be preserved.				
	1				
	827: contig of 827 bp in length				
	gap of unknown length				
	828 1737: contig of 910 bp in length				
	gap of unknown length				
	1738 2655: contig of 918 bp in length				
	gap of unknown length				
	2656 3510: contig of 855 bp in length				
	gap of unknown length				
	3511 4384: contig of 874 bp in length				
	gap of unknown length				
	4385 5298: contig of 914 bp in length				
	gap of unknown length				
	5299 6231: contig of 933 bp in length				
	gap of unknown length				
	6232 7060: contig of 829 bp in length				
	gap of unknown length				
	7061 7952: contig of 892 bp in length				
	gap of unknown length				
	7953 8847: contig of 895 bp in length				
	gap of unknown length				
	8848 9706: contig of 859 bp in length				
	gap of unknown length				
	9707 10550: contig of 844 bp in length				
	gap of unknown length				
	10551 11431: contig of 881 bp in length				
	gap of unknown length				
	11432 gap of unknown length				
	12384: contig of 953 bp in length				
	gap of unknown length				
	13248: contig of 864 bp in length				
	gap of unknown length				
	14083: contig of 835 bp in length				
	gap of unknown length				
	14955: contig of 872 bp in length				
	gap of unknown length				
	15823: contig of 868 bp in length				
	gap of unknown length				
	16675: contig of 852 bp in length				
	gap of unknown length				
	17644: contig of 969 bp in length				
	gap of unknown length				
	18541: contig of 897 bp in length				
	gap of unknown length				
	19391: contig of 850 bp in length				
	gap of unknown length				
	20239: contig of 848 bp in length				
	gap of unknown length				
	21078: contig of 839 bp in length				
	gap of unknown length				
	21897: contig of 819 bp in length				
	gap of unknown length				
	22812: contig of 915 bp in length				
	gap of unknown length				
	23708: contig of 896 bp in length				
	gap of unknown length				
	24563: contig of 855 bp in length				
	gap of unknown length				
	25430: contig of 867 bp in length				
	gap of unknown length				
	26328: contig of 898 bp in length				
	gap of unknown length				
	27131: contig of 803 bp in length				
	gap of unknown length				
	28001: contig of 870 bp in length				
	gap of unknown length				
	28961: contig of 960 bp in length				
	gap of unknown length				
	29913: contig of 952 bp in length				
	gap of unknown length				
	30776: contig of 863 bp in length				
	gap of unknown length				
	31632: contig of 856 bp in length				
	gap of unknown length				
	32525: contig of 893 bp in length				
	gap of unknown length				
	33367: contig of 842 bp in length				
	gap of unknown length				
	34288: contig of 921 bp in length				
	gap of unknown length				
	35097: contig of 809 bp in length				
	gap of unknown length				
	36050: contig of 953 bp in length				
	gap of unknown length				
	36942: contig of 892 bp in length				
	gap of unknown length				
	37783: contig of 841 bp in length				
	gap of unknown length				
	38640: contig of 857 bp in length				
	gap of unknown length				
	39503: contig of 863 bp in length				
	gap of unknown length				
	40441: contig of 938 bp in length				
	gap of unknown length				
	41319: contig of 878 bp in length				
	gap of unknown length				
	42139: contig of 820 bp in length				
	gap of unknown length				
	42883: contig of 744 bp in length				
	gap of unknown length				

* 42884	* 43606:	contig of 723 bp in length	
*		gap of unknown length	
* 43607	* 44330:	contig of 724 bp in length	
*		gap of unknown length	
* 44331	* 45057:	contig of 727 bp in length	
*		gap of unknown length	
* 45058	* 45787:	contig of 730 bp in length	
*		gap of unknown length	
* 45788	* 46513:	contig of 726 bp in length	
*		gap of unknown length	
* 46514	* 47241:	contig of 728 bp in length	
*		gap of unknown length	
* 47242	* 47965:	contig of 724 bp in length	
*		gap of unknown length	
* 47966	* 48688:	contig of 723 bp in length	
*		gap of unknown length	
* 48689	* 49426:	contig of 738 bp in length	
*		gap of unknown length	
* 49427	* 50142:	contig of 716 bp in length	
*		gap of unknown length	
* 50143	* 50889:	contig of 747 bp in length	
*		gap of unknown length	
* 50890	* 51613:	contig of 724 bp in length	
*		gap of unknown length	
* 51614	* 52344:	contig of 731 bp in length	
*		gap of unknown length	
* 52345	* 53080:	contig of 736 bp in length	
*		gap of unknown length	
* 53081	* 53840:	contig of 760 bp in length	
*		gap of unknown length	
* 53841	* 54571:	contig of 731 bp in length	
*		gap of unknown length	
* 54572	* 55300:	contig of 729 bp in length	
*		gap of unknown length	
* 55301	* 56024:	contig of 724 bp in length	
*		gap of unknown length	
* 56025	* 56765:	contig of 741 bp in length	
*		gap of unknown length	
* 56766	* 57489:	contig of 724 bp in length	
*		gap of unknown length	
* 57490	* 58214:	contig of 725 bp in length	
*		gap of unknown length	
* 58215	* 58937:	contig of 723 bp in length	
*		gap of unknown length	
* 58938	* 59658:	contig of 721 bp in length	
*		gap of unknown length	
* 59659	* 60380:	contig of 722 bp in length	
*		gap of unknown length	
* 60381	* 61105:	contig of 725 bp in length	
*		gap of unknown length	
* 61106	* 61830:	contig of 725 bp in length	
*		gap of unknown length	
Query Match 4.3%; Score 47.6; DB 42; Length 230475;			
Best Local Similarity 61.8%; Pred. No 0.0093;			
Matches 94; Conservative 0; Mismatches 54; Indels 4; Gaps 1;			
QY 952	aacagctggcaggtggcacagtttataatctcagcccttgaagctcaggctggaga 1011		
DB 100621	ACCTGCTGGCGGTGGTGGCGCAGCGCTTTAATCCAGCACTTGGGAGCAAGCGAGGTGA 100680		
QY 1012	atgggaagtgaagctggcctgtcttcattagtaggctcagtcga----attaaag 1067		
DB 100681	TTTCTGAGTTTGGAGCGAGCTGGTGTACAGTGTGAGTTCACGACAGCGAGCTACAC 100740		
QY 1068	agtaagaacactattataaaaaa 1099		
DB 100741	AGGAAACCTATCTTGAAACAAACACACACA 100772		

RESULT 13
AC020965/c
LOCUS

AC020965 -238554 bp · DNA HTG 10-FEB-2000

DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
AUTHORS
TITLE
JOURNAL
COMMENT

Mus musculus clone RP21-82116, WORKING DRAFT SEQUENCE, 80 unordered pieces.
AC020965
AC020965.1 GI:6691265
HTG; HTGS_PHASE1; HTGS_DRAFT.
house mouse.
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 238554)
DOE Joint Genome Institute.
Sequencing of Mouse
Unpublished
2 (bases 1 to 238554)
DOE Joint Genome Institute.
Direct Submission
Submitted (12-JAN-2000) Production Sequencing Facility, DOE Joint
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
-----Genome Center
Center: Joint Genome Institute
Center Code: JGI
Web site: http://www.jgi.doe.gov
-----Summary Statistics
Consensus quality: 160205 bases at least Q40
Consensus quality: 219616 bases at least Q30
Consensus quality: 238602 bases at least Q20
Estimated insert size: 238554; sum-of-contigs estimation
Estimated insert size: 162000; pulse field gel estimation
Quality coverage: 4.94x in Q20 bases; pulse field gel estimation
Quality coverage: 3.36x in Q20 bases; sum-of-contigs estimation

* NOTE: This is a 'working draft' sequence. It currently
* consists of 80 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

1 1259: contig of 1253 bp in length
gap of unknown length
1254 2582: contig of 1329 bp in length
gap of unknown length
2583 3769: contig of 1187 bp in length
gap of unknown length
3770 4974: contig of 1205 bp in length
gap of unknown length
4975 6214: contig of 1240 bp in length
gap of unknown length
6215 7484: contig of 1270 bp in length
gap of unknown length
7485 8808: contig of 1324 bp in length
gap of unknown length
8809 10079: contig of 1271 bp in length
gap of unknown length
10080 11473: contig of 1394 bp in length
gap of unknown length
11474 12921: contig of 1448 bp in length
gap of unknown length
12922 14128: contig of 1207 bp in length
gap of unknown length
14129 15531: contig of 1403 bp in length
gap of unknown length
15532 17044: contig of 1513 bp in length
gap of unknown length
17045 18665: contig of 1621 bp in length
gap of unknown length
18666 19891: contig of 1226 bp in length
gap of unknown length
19892 21416: contig of 1525 bp in length
gap of unknown length
21417 22724: contig of 1308 bp in length
gap of unknown length

"

"

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 15, 2000, 01:54:21 ; Search time 1646.42 seconds
(without alignments)
-800.605 Million cell updates/sec

Title: US-09-223-796-3
Perfect score: 1355
Sequence: 1 gggcaggcagtgagggtga.....gtgttcaggcaggcccg 1355

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 882769 seqs, -486395729 residues

Total number of hits satisfying chosen parameters: 1765538

Minimum DB seq length: 0
Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database : GenEmbl.*

- 1: gb_ba1.*
- 2: gb_ba2.*
- 3: gb_ov.*
- 4: gb_ov.*
- 5: gb_pat.*
- 6: gb_ph.*
- 7: gb_pl1.*
- 8: gb_pl2.*
- 9: gb_pr1.*
- 10: gb_pr2.*
- 11: gb_pr3.*
- 12: gb_ro.*
- 13: gb_sts.*
- 14: gb_sy.*
- 15: gb_un.*
- 16: gb_v1.*
- 17: em_fun.*
- 18: em_hum1.*
- 19: em_hum2.*
- 20: em_in.*
- 21: em_cm.*
- 22: em_or.*
- 23: em_ov.*
- 24: em_pat.*
- 25: em_ph.*
- 26: em_pl.*
- 27: em_ro.*
- 28: em_sts.*
- 29: em_sy.*
- 30: em_un.*
- 31: em_v1.*
- 32: gb_htg1.*
- 33: gb_htg2.*
- 34: gb_in1.*
- 35: gb_in2.*
- 36: em_ba1.*
- 37: em_ba2.*
- 38: em_hum3.*
- 39: em_hum4.*
- 40: gb_pr4.*
- 41: gb_htg3.*
- 42: gb_htg4.*
- 43: gb_htg5.*
- 44: gb_htg6.*

- 45: gb_htg7.*
- 46: em_htg1.*
- 47: em_htg2.*
- 48: em_htg3.*
- 49: em_hum5.*
- 50: gb_pl3.*
- 51: gb_pr5.*
- 52: gb_htg8.*
- 53: gb_htg9.*
- 54: gb_htg10.*
- 55: gb_htg11.*
- 56: gb_htg12.*
- 57: gb_htg13.*
- 58: gb_htg14.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	1144.4	84.5	1588	40	AF113540	AF113540 Homo sapi
C 2	1085.8	80.1	108369	40	AF124523	AF124523 Homo sapi
C 3	1084.2	80.0	174923	52	AF216665	AF216665 Homo sapi
4	1081	79.8	128009	40	AF118808	AF118808 Homo sapi
5	1077.8	79.5	69528	40	AF146367	AF146367 Homo sapi
C 6	724.2	53.4	133200	43	AC011951	AC011951 Homo sapi
7	110.8	8.2	347	9	HS79C3R	263154 H.sapiens C
8	104	7.7	1041	40	AF014882	AF014882 Homo sapi
9	104	7.7	1041	40	AF014883	AF014883 Homo sapi
10	104	7.7	1041	40	AF014884	AF014884 Homo sapi
11	104	7.7	1041	40	AF014885	AF014885 Homo sapi
12	104	7.7	1041	40	AF014886	AF014886 Homo sapi
13	104	7.7	1041	40	AF014887	AF014887 Homo sapi
14	104	7.7	1041	40	AF014888	AF014888 Homo sapi
15	104	7.7	1041	40	AF014889	AF014889 Homo sapi
16	104	7.7	1041	40	AF014890	AF014890 Homo sapi
17	104	7.7	1041	40	AF014891	AF014891 Homo sapi
18	104	7.7	1041	40	AF014892	AF014892 Homo sapi
19	104	7.7	1041	40	AF014893	AF014893 Homo sapi
20	104	7.7	1041	40	AF014894	AF014894 Homo sapi
21	104	7.7	1041	40	AF014895	AF014895 Homo sapi
22	104	7.7	1041	40	AF014896	AF014896 Homo sapi
23	104	7.7	1041	40	AF014897	AF014897 Homo sapi
24	104	7.7	1041	40	AF014898	AF014898 Homo sapi
25	104	7.7	1041	40	AF014899	AF014899 Homo sapi
26	104	7.7	1041	40	AF014900	AF014900 Homo sapi
27	104	7.7	15447	10	MTHSGENOM	X62996 H.sapiens m
28	104	7.7	16559	9	HUMMTA	D38112 Homo sapien
29	104	7.7	16570	11	HSMITG	X93334 Homo sapien
30	102.4	7.6	1041	40	AF014901	AF014901 Homo sapi
31	102.4	7.6	1320	40	U95646	U95646 Homo sapien
C 32	102.4	7.6	2771	10	HUMMTM1	M10546 Human mitoc
33	102.4	7.6	16569	10	MHSXX	V00662 H.sapiens m
34	102.4	7.6	16569	51	HUMMTG	J01415 Human mitoc
35	101.8	7.5	5840	40	AF134583	AF134583 Homo sapi
36	91.2	6.7	1041	11	AF014903	AF014903 Pan trogl
37	91.2	6.7	16563	9	CHPMT	D38116 Pan paniscu
38	89.6	6.6	1041	11	AF014904	AF014904 Pan trogl
39	89	6.6	1041	11	AF014902	AF014902 Pan trogl
40	89	6.6	1041	11	AF014905	AF014905 Pan trogl
41	89	6.6	1041	11	AF014907	AF014907 Pan trogl
42	89	6.6	1041	11	AF014911	AF014911 Pan trogl
43	89	6.6	1041	11	AF014915	AF014915 Pan trogl
44	89	6.6	1041	11	AF014919	AF014919 Pan trogl
45	89	6.6	1041	11	AF014921	AF014921 Pan trogl

ALIGNMENTS

RESULT	1
AF113540	1588 bp mRNA PRI 04-DEC-1999
LOCUS	Homo sapiens hypothalamus protein HT002 mRNA, complete cds.
DEFINITION	AF113540 Homo sapiens hypothalamus protein HT002 mRNA, complete cds.
ACCESSION	AF113540.1 GI:6523834
VERSION	
KEYWORDS	human.
SOURCE	Homo sapiens
ORGANISM	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE	1 (bases 1 to 1588)
AUTHORS	Dong,H., Ren,S., Huang,C., Jiang,C., Li,Y., Zhou,J., Yu,Y., Xu,S., Wang,Y., Fu,G., Chen,Z. and Han,Z.
TITLE	A novel gene expressed in the human hypothalamus
JOURNAL	Unpublished
REFERENCE	2 (bases 1 to 1588)
AUTHORS	Dong,H., Ren,S., Huang,C., Jiang,C., Li,Y., Zhou,J., Yu,Y., Xu,S., Wang,Y., Fu,G., Chen,Z. and Han,Z.
TITLE	Direct Submission
JOURNAL	Submitted (16-DEC-1998) Chinese National Human Genome Center at Shanghai, 351 Guo Shoujing Rd., Zhangjiang Hi-Tech Park, Pudong, Shanghai 201203, China
FEATURES	Location/Qualifiers
source	1..1588
	/organism="Homo sapiens"
	/db_xref="taxon:9606"
	/tissue_type="hypothalamus"
CDS	418..1089
	/codon_start=1
	/product="hypothalamus protein HT002"
	/protein_id="AAFL487.1"
	/db_xref="GI:6523835"
	/translation="MSAQTATPYLHHPGDHSGRVSFGLQAQPPEVAAMRLGLDL RSTFRLLKLVSSLOEDRCDCSVLGSAPTCRRAGGILLAGMHTLLQALRLPPTS LKPDTRDPAPGALHPRPGRGLGWIEPAALLDSVQQCGAWLPHVADFWRVDV AISTLSARLSQPSVLMLQLSDGSAYREVPVTKFQLRYSLVALVKEMADLEKRKE RLQD"
BASE COUNT	354 a 437 c 437 g 360 t
ORIGIN	
Query Match	84.5%; Score 1144.4; DB 40; Length 1588;
Best Local Similarity	99.2%; Pred. No. 7.9e-252;
Matches 1231:	Conservative 0; Mismatches 16; Indels 7; Gaps 7;
Qy	100 aaaccccgaggcgcagaccgtccacacgggacaggagcccgcctcgctccgccgctccg 159
Db	224 AAACCTGGAGGCGCAGACCCTCCACACGAGACAGGGACGCCGCCTCGCTCCGTC 283
Qy	160 cgctgatcctcaaggcttcgcgggtttccccgcgcccacccggacgcgcgaagaagcag 219
Db	284 -GCTGATCCTCAAGGTCTCGCGGTTTCCC CGCCGCCACCCGCGACGCCGAGAACGCAG 342
Qy	220 cgagctctcagctcaggaatctgcattctgggacctcttggtggctgcatcaaa 279
Db	343 CGAGCTCTCAGTCTCAGGCATCTGCATCTGGGACCGACCTCCGGGCTGGCTGATCAA 402
Qy	280 gaggaagcagcagaatctctgctgtgggggctgcaactccatcacctgcatactcctggt 339
Db	403 GAGGAGCAGCAGAATCTGCTGTGGGACTGCAACTCATCTGATCATCTCTGGT 462
Qy	340 gatagtacagtgccgagtagttagttcttggggggcccgactctccagaggtggcagca 399
Db	463 GATAGTCACAGTGCCGAGTAGTGTCTTGTGGGGCCCCAGCTTCCTCCAGAGGTGGCAGCA 522
Qy	400 atggcccggtactagggaacccagagcaggttacaaaagtgtcctaagtttgt 459
Db	523 ATGGCCCCGGCTACTAGGGGACCTACACAGGAGCACGTTTCAGAAAAGTTCTGAAGTTGTG 582
Qy	460 gtccagcagctgcagggggagactgccgagacggtgtgcagcgctcttgggggtcagcgcc 519

Merck-Roussseau, M.F. and Rosenthal, A.
Direct Submission
Submitted (29-JAN-1999) Genome Analysis, Institute of Molecular
Biotechnology, Beutenbergstrasse 11, Jena 07745, Germany
FEATURES
 source
 1. .108369
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="8"
 /clone="Kox18-21"
 1. .245
 /rpt_family="AluY"
 /evidence=not_experimental
 246. .584
 /rpt_family="L1MA4"
 /evidence=not_experimental
 585. .863
 /rpt_family="AluJo"
 /evidence=not_experimental
 864. .1068
 /rpt_family="L1MA4"
 /evidence=not_experimental
 1069. .1353
 /rpt_family="AluSx"
 /evidence=not_experimental
 1387. .1734
 /rpt_family="L1MA4"
 /evidence=not_experimental
 1735. .2042
 /rpt_family="AluY"
 /evidence=not_experimental
 2043. .2172
 /rpt_family="L1MA4"
 /evidence=not_experimental
 2173. .2482
 /rpt_family="AluSx"
 /evidence=not_experimental
 2483. .2901
 /rpt_family="L1MA4"
 /evidence=not_experimental
 2835. .2912
 /note="GRAIL, score = 45.000%, comment = marginal"
 2848. .2883
 /note="Xpound exon prediction, score = 77% (0%)"
 /evidence=not_experimental
 2902. .3202
 /rpt_family="AluSp"
 /evidence=not_experimental
 3203. .3325
 /rpt_family="L1MA4"
 /evidence=not_experimental
 3326. .3625
 /rpt_family="AluSx"
 /evidence=not_experimental
 3626. .3968
 /rpt_family="L1MA4"
 /evidence=not_experimental
 4104. .4431
 /rpt_family="L2"
 /evidence=not_experimental
 4123. .4369
 /note="GRAIL, score = 63.000%, comment = good"
 /evidence=not_experimental
 4450. .4608
 /rpt_family="AluSg/x"
 /evidence=not_experimental
 4623. .4796
 /rpt_family="L2"
 /evidence=not_experimental
 4898. .4997
 /rpt_family="AluSg"
 /evidence=not_experimental

5000. .5060
 /rpt_family="L1MA4A"
 /evidence=not_experimental
5101. .5402
 /rpt_family="AluY"
 /evidence=not_experimental
 complement(5578. .5595)
 /rpt_family="MER33"
 /evidence=not_experimental
5596. .5867
 /rpt_family="AluSx"
 /evidence=not_experimental
5902. .5972
 /rpt_family="AluSx"
 /evidence=not_experimental
 complement(5913. .6139)
 /rpt_family="MER33"
 /evidence=not_experimental
 complement(8764. .9075)
 /rpt_family="AluSx"
 /evidence=not_experimental
10935. .11017
 /rpt_family="MER74B"
 /evidence=not_experimental
 complement(11018. .11306)
 /rpt_family="AluJb"
 /evidence=not_experimental
11307. .11724
 /rpt_family="MER74B"
 /evidence=not_experimental
11725. .12006
 /rpt_family="AluSg"
 /evidence=not_experimental
12049. .12100
 /note="homology = 100.00%, score = 26, counts = 2"
 /rpt_type=tandem
 /rpt_unit=actcacttgatgtgggtgtctgc
 /evidence=not_experimental
 complement(12081. .12230)
 /note="GRAIL, score = 56.000%, comment = good"
 /evidence=not_experimental
 complement(12255. .12748)
 /rpt_family="L1R42"
 /evidence=not_experimental
14255. .14377
 /note="Genscan, score = 10.27%, comment = Initial_exon 123
 bp frame: 1 phase: 0"
 /evidence=not_experimental
14386. .14580
 /gene="zinc-finger protein 7 (ZFP7)"
 /number=1
 join(14386. .14580,15811. .15858,16246. .16372,23346. .23462,
 27285. .29150)
 /gene="zinc-finger protein 7 (ZFP7)"
 /note="mRNA acc.no. M29580"
 /product="ZFP7"
 14386. .29150
 /gene="zinc-finger protein 7 (ZFP7)"
 15176. .15595
 /rpt_family="L2"
 /evidence=not_experimental
15811. .15858
 /gene="zinc-finger protein 7 (ZFP7)"
 /number=2
 join(15856. .15858,16246. .16372,23346. .23462,27285. .29098)
 /gene="zinc-finger protein 7 (ZFP7)"
 /note="mRNA acc.no. M29580"
 /codon_start=1
 /product="zinc-finger protein 7"
 /protein_id="AADI7326.1"
 /db_xref="GI:4325310"
 /translation="MEVTFGDVAHFSSREWCQCLDPQALYREVLEHSSVAGLA
 GFLVFKPELISRLPQSGEPWYLDLQGAEGTEAPRTSKTDTSTIRTEQACEDMDILKS

gap of unknown size: 128689-134145: contig of 5457 bp;
 134145-134146: gap of unknown size; 134146-147620: contig of 13475
 bp; 147620-147621: gap of unknown size; 147621-149478: contig of
 1858 bp; 149478-149479: gap of unknown size; 149479-151777:
 contig of 2299 bp; 151777-151778: gap of unknown size;
 151778-151779: contig of 3353 bp; 151779-151780: gap of unknown
 size; 151780-151781: contig of 2124 bp; 151781-151782: gap of
 unknown size; 151782-151783: contig of 1238 bp; 151783-151784:
 gap of unknown size; 151784-151785: contig of 16431 bp.
 * NOTE: This is a 'working draft' sequence.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.

FEATURES

source	Location/Qualifiers
	1. 174923
	/organism="Homo sapiens"
	/db_xref="taxon:9606"
	/chromosome="8"
	/map="8q24"
	/clone="BAC 704B241056"
	complement(36..349)
repeat_region	/rpt_family="AluX"
repeat_region	352..405
exon	/rpt_type=tandem
	complement(449..582)
	/note="FEXHB"
repeat_region	/evidence=not_experimental
	complement(457..748)
repeat_region	/rpt_family="AluSg"
	complement(750..820)
repeat_region	/rpt_family="L1M4_orf2"
	complement(1167..1754)
repeat_region	/rpt_family="L1M4_orf2"
	1387..1410
repeat_region	/rpt_type=tandem
	complement(1410..1713)
exon	/rpt_family="AluY"
	1434..1555
	/note="FEXHB"
repeat_region	/evidence=not_experimental
	1517..1702
repeat_region	/rpt_type=inverted
	complement(1865..2170)
repeat_region	/rpt_family="AluJo"
	2386..2461
	/rpt_type=tandem
repeat_region	complement(2435..2544)
repeat_region	/rpt_family="AluJb"
	complement(2545..2649)
repeat_region	/rpt_family="AluSx"
	2738..3017
repeat_region	/rpt_family="AluJo"
	2747..2931
exon	/rpt_type=inverted
	3369..3458
exon	/note="GENSCAN"
	4066..4182
exon	/note="GENSCAN"
	4066..4182
	/note="MZF"
repeat_region	/evidence=not_experimental
	complement(4603..5038)
exon	/rpt_family="MLTIC"
	complement(4836..4889)
	/note="GRAIL"
exon	/evidence=not_experimental
	complement(5199..5235)
	/note="GRAIL"
exon	/evidence=not_experimental
	complement(5199..5235)
	/note="MZF"
repeat_region	/evidence=not_experimental
	5336..5612

repeat_region	/rpt_family="AluY"
	5472..5530
repeat_region	/rpt_type=inverted
	5627..5906
repeat_region	/rpt_type=tandem
	complement(6589..6751)
repeat_region	/rpt_family="AluSg"
	6689..6747
exon	/rpt_type=inverted
	8005..8268
	/note="MZF"
exon	/evidence=not_experimental
	8005..9818
exon	/note="GENSCAN"
	8046..9818
	/note="FEXHB"
exon	/evidence=not_experimental
	8046..9818
	/note="GRAIL"
exon	/evidence=not_experimental
	complement(9207..9412)
	/note="GRAIL"
repeat_region	/evidence=not_experimental
	9299..9578
repeat_region	/rpt_type=tandem
	9558..10259
exon	/rpt_family="AluY"
	complement(10113..10232)
	/note="FEXHB"
repeat_region	/evidence=not_experimental
	10238..10259
repeat_region	/rpt_type=tandem
	11078..11384
repeat_region	/rpt_family="AluY"
	11359..11384
repeat_region	/rpt_type=tandem
	11519..11638
repeat_region	/rpt_type=tandem
	complement(11792..11848)
exon	/rpt_family="L1R40b"
	complement(11834..11860)
	/note="GENSCAN"
repeat_region	complement(12208..12332)
exon	/rpt_family="HERVL40"
	complement(12368..12424)
exon	/note="GENSCAN"
	complement(12368..12424)
	/note="XPOUND"
exon	/evidence=not_experimental
	complement(12368..12468)
	/note="GRAIL"
exon	/evidence=not_experimental
	complement(12673..12773)
	/note="GRAIL"
repeat_region	/evidence=not_experimental
	12872..12961
repeat_region	/rpt_family="HERVL40"
	12985..13069
repeat_region	/rpt_family="MER58B"
	13354..13491
repeat_region	/rpt_type=inverted
	complement(13354..13641)
repeat_region	/rpt_family="AluY"
	13949..14111
repeat_region	/rpt_family="FRAM"
	14263..14357
exon	/rpt_family="L2b"
	complement(14534..14650)
	/note="GRAIL"
repeat_region	/evidence=not_experimental
	14715..14969
repeat_region	/rpt_family="AluJo"
	15002..15162


```
/note="Exon predicted by several exon prediction programs
and by homology to GenBank entry HUM2F1A.1: Homo sapiens
zinc finger protein mRNA, 3' end."
/evidence=not_experimental
2269..2310
exon

/note="Exon predicted by several exon prediction programs
and by homology to GenBank entry HUM2NFN.1: Homo sapiens
DNA-binding protein (ZNF) gene, partial cds."
/evidence=not_experimental
47097..47263
exon

/note="Exon predicted by several exon prediction programs
and by homology to GenBank entry HSKOX9.1: Human Kox9 mRNA
for zinc finger protein, partial."
/evidence=not_experimental
47430..47631
exon

/note="Exon predicted by several exon prediction programs
and by homology to GenBank entry AF024700.1: Homo sapiens
clone BA-2 zinc finger protein mRNA, partial cds."
/evidence=not_experimental
94802..95455
exon

/note="Exon predicted by several exon prediction programs
and by homology to GenBank entry HS2FP647.1: Human mRNA
for zinc finger protein (clone 647)."
/evidence=not_experimental
complement(95143..95187)
exon

/note="Exon predicted by several exon prediction programs
and by homology to GenBank entry HS2NFP7.1: H. sapiens
mRNA for zinc-finger protein (ZNFp7)."
/evidence=not_experimental
BASE COUNT 35117 a 30514 c 28979 g 33352 t 47 others
ORIGIN

Query Match 79.8%; Score 1081; DB 40; Length 128009;
Best Local Similarity 98.6%; Pred. No. 2.6e-237;
Matches 1101; Conservative 0; Mismatches 15; Indels 1; Gaps 1;

2; 236 aggcattcgcattcgggaccacacctctcctcggctggtgatacagcagca 295
Db 125842 AGGCATCTGCATCTGGGACCCACCTCTGGCTGGCTGATCAAGAGAGACAGCA 125901

Qy 296 tgcctgctggtgggctgcaactcaccctgcatcctcgtggtgatacagcagtgccc 355
Db 125902 TGCTCTGCTGGGGCTGCACACTCCATCCTGCATCTCCTGGTGTGATGCACAGTGGCC 125961

Qy 356 gagtgcgttttggggccagcttctccagaggtggcagcaatggcccggtactag 415
Db 125962 GAGTGAGTTCCTTGGGGGCCAGCTTCTCCAAAGGTGGCAGCAATGGCCGGCTACTAG 126021

Qy 416 gggacctagacagacagcttcagaaagtgtcgaagtgtggtcagcagcgtcgagg 475
Db 126022 GGGACCTAGACAGAGCAGCTTCAGAAAGTGTCTGAAGTGTGTGTCAGCAGCGCTGCAGG 126081

Qy 476 gggaggactgcggagacggtgtgagcgtcttggggtcagcgcacacactccagagagc 535
Db 126082 GGGAGGACTGCCGAAGGCTGTGCACCTGTGGGGTTCAGGCCAACCTGCCGGAGGAGC 126141

Qy 536 agctgggtgcctgctgctggcaggcatgcacacactgctccagcagggccctcgtctgccc 595
Db 126142 AGCTGGGTGCTCCTGCTGGCAGGATGCACACTGCTCCAGCAGGCGCTCCGCTGCCCC 126201

Qy 596 ccaccagcctgaagcctgcacacttcaggacacagctccaggagcctgtgaccccccaag 655
Db 126202 CCACCAGCCTGAAGCCTTGACACTTTCAGGAGCAGCAGCTCCAGGAGCTGTGATCCCCAAG 126261

Qy 656 acctggtcggggacttggcagcgtgtgatttgggagcagcagggccctcctgtgattctg 715
Db 126262 ACCTGGTCGGGACTTGGCCAGCGTGTATTGTTGGAGCCAGCGGCCCTCTTGTATCTG 126321

Qy 716 tggccagcagcaggggacctggtgcgcagcatgttgcgacttccgtggtgggtggatg 775
Db 126322 TGCCCCAGCAGCAGGGGGCCCTGGCTGCCGATCTGCTGACTTTCGTTGGCGGTGGATG 126391
```

```
Qy 776 tagcaattccaccagtgccctggctgcctccctgcagccgagcgtcctgatgcagctga 835
Db 126382 TAGCAATCTCCACAGTGCCTTGGCTGCCTCCCTGCAGCCGAGCGTCTCTGATCAGCTGA 126441

Qy 836 agcttcagatgggtcagcaccgctttgagttcccccacagccaaagttccagagctgc 895
Db 126442 AGCTTTCAGATGGGTGACATACCGCTTTGAGGTCCCCACAGCCAAAGTTCACGAGAGTGC 126501

Qy 896 ggtacagcgtggccctggtcctaaaggagatggcagatcctggagaaagctgtgagcgca 955
Db 126502 GGTACAGCGTGGCCCTGGTCTCTAAAGGAGATGGCAGATCTGGAGAGAGAGTGTGAGCGCA 126561

Qy 956 gactgcaggactgaccctcacttgaccagctccattccagctcggcttgacagcagcc 1015
Db 126562 GACTGCAGGACTGACCCCTCCTCCTGACCTGACCTCCATTCAGATCCGGCTTGGACAGGCACC 126621

Qy 1016 tgagatggttccaaagtgcagctgactcttcccacacacagcctggccttcccacagcc 1075
Db 126622 TGAGATGGTCCAAAGTGCAGCTGACTTCCCCACACAGCCCTGCCCTTCCCATGAGGC 126681

Qy 1076 agccttcagtgagtggttgaacgtaattatgtagtttctgttttaattgaaaaagaga 1135
Db 126682 AGCCTCTTCAGTGAGTGTGTAACGTAATTATGATGTTTCTGTTTAATTGAAAAGAGA 126741

Qy 1136 gctatgccttttttcttttggaaagtaaaagcagct-aaacatgtttctatatagtgagt 1194
Db 126742 GCTATGCCCTTTTCTTTTGGAAAGTAAGCAGCTAAAAACATGTTTCTATAGTGTAGT 126801

Qy 1195 gttggaccttcacacctcccttccctgtacatttcttggctgacgtggcagctg 1254
Db 126802 GTTGGACCTCACACCTCCCTTCCCTGTCATTTGTTGCTGGTGGAGTGGCCATG 126861

Qy 1255 tgaggccaggttgaggccctttgtagacaacatacacagcttgctcagcctggcccatgtag 1314
Db 126862 TGAGGGCAGGTTGAGGCCCTTTGTAGACAACATACAGTGTGCTCAGCCTGGCCCCATGTAG 126921

Qy 1315 ceaggtgctttttagatcttctgttttcaggcaggc 1351
Db 126922 CCAGGTGCTTTGTAGATCTTGTTGTTTAGGTTGGGC 126958

RESULT 5
AF146367
LOCUS
DEFINITION Homo sapiens chromosome 8 PAC P584.00 containing gene for zinc
finger protein, complete sequence.
ACCESSION AF146367
VERSION AF146367.1 GI:4877991
KEYWORDS HTG.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 69528)
AUTHORS Rump,A., Schudy,A., Drescher,B., Koczan,D., Thiessen,H.-J.,
Merck-Roussseau,M.-F. and Rosenthal,A.
Direct Submission
JOURNAL Submitted (28-APR-1999) Genome Analysis, Institute of Molecular
Biotechnology, Beutenbergstrasse 11, Jena 07745, Germany
FEATURES
Location/Qualifiers
1..69528
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="8"
/clone="PAC P584.00"
5504..6968
/gene="hypothetical"
/note="GenScan, score = 1186%, comment = PL_377_aa 1..208
of owlP156221ZF64_HUMAN ZINC FINGER PROTEIN CLONE 647
(FRAGMENT). - HOMO SAPIENS P = 1.2e-120 S = 1186"
/evidence=not_experimental
5504..6637
/gene="hypothetical"
```



```

QY 1195 gttgacattcacacctccctccctgtacattgtcttggtagcagtgccatg 1254
|||||
Db 37715 GTTGACCTCACACCTCCCTTCCTGTACATTTGTTTGGTGCTGAGAGTGCCCATG 37774
*
QY 1255 ttagccagggtgagcccttctgtgagacaacatacagttgtcagctggcccatgtag 1314
|||||
Db 37775 TGAGGCCAGGTGAGGCCCTTGTGTAGACAACATACAGTTGCTACGCTGGCCCATGTAG 37834
*
QY 1315 ccagtgctttgtgatctgtgtgttcagcaggcc 1351
|||||
Db 37835 CCAGTGCTTTGTAAATCTGTGTTTGTAGTTGGC 37871
*
RESULT 6
AC011951/c
LOCUS AC011951 133200 bp DNA HTG 13-DEC-1999
DEFINITION Homo sapiens clone RP11-18M17, *** SEQUENCING IN PROGRESS ***, 53
unordered pieces.
ACCESSION AC011951
VERSION AC011951.2 GI:6554541
KEYWORDS HTG; HTGS_PHASE1.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 133200)
Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,
Baldwin,J., Barna,N., Beckerly,R., Boguslavsky,L., Boukhgalter,B.,
Brown,A., Castle,A., Colangelo,M., Collins,S., Collymore,A.,
Cooke,P., Dearellano,K., Dewar,K., Domino,M., Donelan,L., Doyle,M.,
Ferreira,P., FitzHugh,W., Forrest,C., Funke,R., Gage,D., Gage,D.,
Galagan,J., Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L.,
Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,
Lehoczky,J., Lieu,C., Locke,K., Macdonald,P., Marquis,N.,
McEwan,P., McGurk,A., McKernan,K., McLaughlin,J., Meldrum,J.,
Morrow,J., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,J.,
Peterson,K., Pollara,V., Riley,R., Roy,A., Santos,R., Severy,P.,
Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
Tesfaye,S., Tirrell,A., Vassiliev,H., Vo,A., Wheeler,J., Wu,X.,
Wyman,D., Ye,W.J., Zimmer,A. and Zody,M.
Direct Submission
Submitted (16-OCT-1999) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Dec 10, 1999 this sequence version replaced gi:6056232.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Web site: http://www-seq.wi.mit.edu
Contact: sequence.submissions@genome.wi.mit.edu
----- Project Information
Center project name: L3701
Center clone name: 18_M_17
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 53 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1 1041: contig of 1041 bp in length
*
* 1042 . 1130: contig of unknown length
*
* 1042 . 1130: contig of 89 bp in length
*
* 1042 . 1130: gap of unknown length
*

```

[illegible]

```

QY 100 acaacccggagcgacacccgtccacacggagacggagcccgctcgctccgctccg 159
Db 127 AANACCCGTGGAGCGCAGNCCGTCACACGGAGGAGGAGCGCCGCTCGGTCCGCTCCG 186
QY 160 cgtatctctcaagtcctcgagg-tttcccgccgcccacccggagcgccgacg-aaagcc 217
Db 187 CGCTGATCTCAAGTCTCGGGTTTCCCGTCGCCACCGGAGCGCCNAGNAAGCC 246
QY 218 acgagctctcagctcagcgacatctgcatctggagcagcactcctggctggc 271
Db 247 ACGGAGCTCCTCAGCCTCAGGTGACTGGAATCTGACCCACACACGGGACGC 300

RESULT 8
AF014882 LOCUS AF014882 1041 bp DNA PRI 06-MAY-1999
DEFINITION Homo sapiens NADH dehydrogenase subunit 2 (ND2) gene, mitochondrial
ACCESSION AF014882
VERSION AF014882.1 GI:2582056
KEYWORDS
SOURCE human.
ORGANISM Mitochondrion Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 1041)
AUTHORS Wise,C.A., Srani,M. and Easteal,S.
TITLE Departure from Neutrality at the Mitochondrial NADH dehydrogenase
subunit 2 gene in Humans, but not in Chimpanzees
JOURNAL Genetics (1997) In press
REFERENCE 2 (bases 1 to 1041)
AUTHORS Wise,C.A.
TITLE Direct Submission
JOURNAL Submitted (21-JUL-1997) Human Genetics Group, John Curtin School of
Medical Research, The Australian National University, Mills Rd,
Canberra, ACT 0200, Australia
FEATURES
source
1. .1041
/organism="Homo sapiens"
/organelle="mitochondrion"
/strain="Afri"
/db_xref="taxon:9606"
/notes="Bantu from Durban, South Africa"
1. .>1041
/gene="ND2"
1. .>1041
/gene="ND2"
/codon_start=1
/transl_table=2
/product="NADH dehydrogenase subunit 2"
/protein_id="AAC25441.1"
/db_xref="GI:2582057"
/translation="MNPLAQPVYISTFAGLTITLSSHWFTWVLENNMLAFIPVL
TKMNPSTEAARKYFLTOATSMILLMAILFNMLSGQWTNTNQYSLIMMMAM
AKLGMAPFHWVPEVTOGTPLTSGLLLTWQKLAPISIMYOISPLSNVSLTLISIL
SIMAGSWGGLNQTLRKILAYSISITHMGMMAVLPYNPMITILNIIYILTTAFLL
LNLSSTTLLSRTNKLWLTPLIPSTLLSLGGLPLTGLPKWAIIEFTKNSL
LIPTMATITLLNLYFYLRILYISTITLLPMSNNVKKWQFEHTKPTPLPLIALTI
LLLPISPFMLIL"
BASE COUNT 325 a 349 c 100 g 267 t
ORIGIN

Query Match 7.7%; Score 104; DB 40; Length 1041;
Best Local Similarity 100.0%; Pred. No. 9e-14;
Matches 104; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 gcaggcagttgagtgattaaacacccagcagctacgaaatcttagcatactctca 62
Db 490 GCAGGAGTTGAGTGATTAAACCAACCCAGCAGTACGAAAATCTTAGCATACTCTCA 549
QY 63 attaccacataggatgaataatagcagttctaccgtacaaccc 106

Query Match 7.7%; Score 104; DB 40; Length 1041;
Best Local Similarity 100.0%; Pred. No. 9e-14;
Matches 104; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

Db 550 ATTACCCACATAGATGAATAATAGCAGTTCTACCGTACAACCC 593

RESULT 9
AF014883 LOCUS AF014883 1041 bp DNA PRI 06-MAY-1999
DEFINITION Homo sapiens NADH dehydrogenase subunit 2 (ND2) gene, mitochondrial
ACCESSION AF014883
VERSION AF014883.1 GI:3287310
KEYWORDS
SOURCE human.
ORGANISM Mitochondrion Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 1041)
AUTHORS Wise,C.A., Srani,M. and Easteal,S.
TITLE Departure from Neutrality at the Mitochondrial NADH dehydrogenase
subunit 2 gene in Humans, but not in Chimpanzees
JOURNAL Genetics (1997) In press
REFERENCE 2 (bases 1 to 1041)
AUTHORS Wise,C.A.
TITLE Direct Submission
JOURNAL Submitted (21-JUL-1997) Human Genetics Group, John Curtin School of
Medical Research, The Australian National University, Mills Rd,
Canberra, ACT 0200, Australia
COMMENT On Jul 2, 1998 this sequence version replaced gi:2582058.
FEATURES
source
1. .1041
/organism="Homo sapiens"
/organelle="mitochondrion"
/strain="Afri"
/db_xref="taxon:9606"
/notes="Bantu from Durban, South Africa"
1. .>1041
/gene="ND2"
1. .>1041
/gene="ND2"
/codon_start=1
/transl_table=2
/product="NADH dehydrogenase subunit 2"
/protein_id="AAC25442.1"
/db_xref="GI:3287311"
/translation="MNPLAQPVYISTFAGLTITLSSHWFTWVLENNMLAFIPVL
TKMNPSTEAARKYFLTOATSMILLMAILFNMLSGQWTNTNQYSLIMMMAM
AKLGMAPFHWVPEVTOGTPLTSGLLLTWQKLAPISIMYOISPLSNVSLTLISIL
SIMAGSWGGLNQTLRKILAYSISITHMGMMAVLPYNPMITILNIIYILTTAFLL
LNLSSTTLLSRTNKLWLTPLIPSTLLSLGGLPLTGLPKWAIIEFTKNSL
LIPTMATITLLNLYFYLRILYISTITLLPMSNNVKKWQFEHTKPTPLPLIALTI
LLLPISPFMLIL"
BASE COUNT 325 a 350 c 100 g 266 t
ORIGIN

Query Match 7.7%; Score 104; DB 40; Length 1041;
Best Local Similarity 100.0%; Pred. No. 9e-14;
Matches 104; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 gcaggcagttgagtgattaaacacccagcagctacgaaatcttagcatactctca 62
Db 490 GCAGGAGTTGAGTGATTAAACCAACCCAGCAGTACGAAAATCTTAGCATACTCTCA 549
QY 63 attaccacataggatgaataatagcagttctaccgtacaaccc 106

Query Match 7.7%; Score 104; DB 40; Length 1041;
Best Local Similarity 100.0%; Pred. No. 9e-14;
Matches 104; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```
ACCESSION AF014884
VERSION AF014884.2 GI:4755089
KEYWORDS
SOURCE human.
ORGANISM Mitochondrion Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
AUTHORS Eutheria; Primates; Catarrhini; Hominidae; Homo.
TITLE 1 (bases 1 to 1041)
JOURNAL Wise, C.A., Sraml, M. and Easteal, S.
REFERENCE Departure from Neutrality at the Mitochondrial NADH dehydrogenase
AUTHORS subunit 2 gene in Humans, but not in Chimpanzees
TITLE 2 (bases 1 to 1041)
JOURNAL Wise, C.A.
REFERENCE Direct Submission
AUTHORS Submitted (21-JUL-1997) Human Genetics Group, John Curtin School of
TITLE Medical Research, The Australian National University, Mills Rd,
JOURNAL Canberra, ACT 0200, Australia
COMMENT On May 6, 1999 this sequence version replaced gi:3287312.
FEATURES
source
1. 1041
/organism="Homo sapiens"
/organelle="mitochondrion"
/strain="Afr3"
/db_xref="taxon:9606"
/note="Bantu from Durban, South Africa"
1. >1041
/gene="ND2"
1. >1041
/gene="ND2"
/transl_table=2
/product="NADH dehydrogenase subunit 2"
/db_xref="GI:2582061"
/translation="MNPLAQPVYISTIFAGTILALSSHWFTWGLENNLAFIPVL
TKMNPSTEAAYKFLQATASMLLMAIFNNLSGQWTMTNTNOYSSLMIMAM
AMKLGAPFHFVPTQGTPLTSGLLLTWKLAPISIMYQISPLNYSLLTSL
SINAGSWGGLNOTLKRILAYSSITHGMGMMAVLPYNPMTILNLTIIITTAFL
LNLSSTTLLSRWNKLTWLTPLIPSTLLSLGGLPPLTGLPKWAIIEFTKNSL
IIPTMATITLLNLYFLRLIYSTISITLLPMSNNVKKMQFEHTKPTPLTIALTT
LLLPISPFMLML"
BASE COUNT 325 a 349 c 100 g 267 t
ORIGIN
Query Match 7.7%; Score 104; DB 40; Length 1041;
Best Local Similarity 100.0%; Pred. No. 9e-14; Indels 0; Gaps 0;
Matches 104; Conservative 0; Mismatches 0;
QY 3 gcaggcagtgagtgattaaacacacacagcagcagcaaatcttagcatactctca 62
|||||
DB 490 GCAGGCAGTGGAGTGATTAACCAACCCAGCTACGCAAAATCTTAGCATACTCTCA 549
|||||
QY 63 attaccacacatagatgaataatagcaggtctaccgtacaacc 106
|||||
DB 550 ATTACCCACATAGGATGAATAATAGCAGTCTTACCGTACAACCC 593
|||||
RESULT 12
AF014886 1041 bp DNA PRI 06-MAY-1999
LOCUS Homo sapiens NADH dehydrogenase subunit 2 (ND2) gene, mitochondrial
DEFINITION gene encoding mitochondrial protein, partial cds.
ACCESSION AF014886
VERSION AF014886.1 GI:3287313
KEYWORDS
SOURCE human.
ORGANISM Mitochondrion Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
AUTHORS Eutheria; Primates; Catarrhini; Hominidae; Homo.
TITLE 1 (bases 1 to 1041)
JOURNAL Wise, C.A., Sraml, M. and Easteal, S.
REFERENCE Departure from Neutrality at the Mitochondrial NADH dehydrogenase
AUTHORS subunit 2 gene in Humans, but not in Chimpanzees
TITLE 2 (bases 1 to 1041)
JOURNAL Wise, C.A.
REFERENCE Direct Submission
AUTHORS Submitted (21-JUL-1997) Human Genetics Group, John Curtin School of
TITLE Medical Research, The Australian National University, Mills Rd,
JOURNAL Canberra, ACT 0200, Australia
COMMENT On Jul 2, 1998 this sequence version replaced gi:2582064.
ACCESSION AF014884
VERSION AF014884.2 GI:4755089
KEYWORDS
SOURCE human.
ORGANISM Mitochondrion Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
AUTHORS Eutheria; Primates; Catarrhini; Hominidae; Homo.
TITLE 1 (bases 1 to 1041)
JOURNAL Wise, C.A., Sraml, M. and Easteal, S.
REFERENCE Departure from Neutrality at the Mitochondrial NADH dehydrogenase
AUTHORS subunit 2 gene in Humans, but not in Chimpanzees
TITLE 2 (bases 1 to 1041)
JOURNAL Wise, C.A.
REFERENCE Direct Submission
AUTHORS Submitted (21-JUL-1997) Human Genetics Group, John Curtin School of
TITLE Medical Research, The Australian National University, Mills Rd,
JOURNAL Canberra, ACT 0200, Australia
COMMENT On May 6, 1999 this sequence version replaced gi:3287312.
FEATURES
source
1. 1041
/organism="Homo sapiens"
/organelle="mitochondrion"
/strain="Afr3"
/db_xref="taxon:9606"
/note="Bantu from Durban, South Africa"
1. >1041
/gene="ND2"
1. >1041
/gene="ND2"
/transl_table=2
/product="NADH dehydrogenase subunit 2"
/db_xref="GI:2582061"
/translation="MNPLAQPVYISTIFAGTILALSSHWFTWGLENNLAFIPVL
TKMNPSTEAAYKFLQATASMLLMAIFNNLSGQWTMTNTNOYSSLMIMAM
AMKLGAPFHFVPTQGTPLTSGLLLTWKLAPISIMYQISPLNYSLLTSL
SINAGSWGGLNOTLKRILAYSSITHGMGMMAVLPYNPMTILNLTIIITTAFL
LNLSSTTLLSRWNKLTWLTPLIPSTLLSLGGLPPLTGLPKWAIIEFTKNSL
IIPTMATITLLNLYFLRLIYSTISITLLPMSNNVKKMQFEHTKPTPLTIALTT
LLLPISPFMLML"
BASE COUNT 325 a 349 c 100 g 267 t
ORIGIN
Query Match 7.7%; Score 104; DB 40; Length 1041;
Best Local Similarity 100.0%; Pred. No. 9e-14; Indels 0; Gaps 0;
Matches 104; Conservative 0; Mismatches 0;
QY 3 gcaggcagtgagtgattaaacacacacagcagcagcaaatcttagcatactctca 62
|||||
DB 490 GCAGGCAGTGGAGTGATTAACCAACCCAGCTACGCAAAATCTTAGCATACTCTCA 549
|||||
QY 63 attaccacacatagatgaataatagcaggtctaccgtacaacc 106
|||||
DB 550 ATTACCCACATAGGATGAATAATAGCAGTCTTACCGTACAACCC 593
|||||
RESULT 11
AF014885 1041 bp DNA PRI 06-MAY-1999
LOCUS Homo sapiens NADH dehydrogenase subunit 2 (ND2) gene, mitochondrial
DEFINITION gene encoding mitochondrial protein, partial cds.
ACCESSION AF014885
VERSION AF014885.1 GI:2582062
KEYWORDS
SOURCE human.
ORGANISM Mitochondrion Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
AUTHORS Eutheria; Primates; Catarrhini; Hominidae; Homo.
TITLE 1 (bases 1 to 1041)
JOURNAL Wise, C.A., Sraml, M. and Easteal, S.
```

```
FEATURES
  source
    Location/Qualifiers
      1..1041
        /organism="Homo sapiens"
        /organelle="mitochondrion"
        /strain="Afr5"
        /db_xref="taxon:9606"
        /note="Bantu from Durban, South Africa"
      1..>1041
        /gene="ND2"
      1..>1041
        /gene="ND2"
        /codon_start=1
        /transl_table=2
        /product="NADH dehydrogenase subunit 2"
        /protein_id="AAC25445.1"
        /db_xref="GI:3287314"
        /translation="MNPQAQVYISTIFAGTLTALSSHWFTWVGLMNMIAFIPVL
        TKMNPSTEAARKYFLQTATSMILLMAILFNMLSGOWTNTTNOYSSLMNMAM
        AMKLGAPFHFVPEVTQGTPLTSGLLLTWQKLAPISIMYQISPSLVSLTLTSL
        SIMAGSWGCGNOTQRLKILAYSSITHMGWMAVLDPYNNPNTILNLYIILTTAFLL
        LNLNSTTLLSRTWKNLTWLTPLIPSLTSLGGLPLTGLPKWAIIEETKKNLSL
        IIPTIMATILNLNLYFLRLIYSTISITLLPMSNVKMKWQFEHTKPTPLLIALT
        LLLPISPFMLMIT"
      325 a 350 c 100 g 266 t
      ORIGIN

Query Match 7.7%; Score 104; DB 40; Length 1041;
Best Local Similarity 100.0%; Pred. No. 9e-14;
Matches 104; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 gcaggcagtgagggtgattaaacacacacagctacgcaaaatcttagcatactctca 62
Db 490 GCAGGCAGTGTGAGTGAATAACCAACCCAGCTACGCAAAATCTTAGCATACTCTCA 549

QY 63 attaccacatagatgaataacacacacagctctaccgtacacccc 106
Db 550 ATTACCCACATAGGATGAATAATAGCAGTCTACCGTACACCC 593

RESULT 13
AF014887
LOCUS AF014887 1041 bp DNA PRI 06-MAY-1999
DEFINITION Homo sapiens NADH dehydrogenase subunit 2 (ND2) gene, mitochondrial
ACCESSION AF014887
VERSION AF014887.1 GI:2582066
KEYWORDS human.
SOURCE Mitochondrion Homo sapiens
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 1041)
AUTHORS Wise,C.A., Sraml,M. and Easteal,S.
TITLE Departure from Neutrality at the Mitochondrial NADH dehydrogenase
subunit 2 gene in Humans, but not in Chimpanzees
JOURNAL Genetics (1997) In press
REFERENCE 2 (bases 1 to 1041)
AUTHORS Wise,C.A.
TITLE Direct Submission
JOURNAL Submitted (21-JUL-1997) Human Genetics Group, John Curtin School of
Medical Research, The Australian National University, Mills Rd,
Canberra, ACT 0200, Australia
COMMENT On Jul 2, 1998 this sequence version replaced gi:2582068.
FEATURES
  source
    Location/Qualifiers
      1..1041
        /organism="Homo sapiens"
        /organelle="mitochondrion"
        /strain="Eur1"
        /db_xref="taxon:9606"
        /note="European: Anglo Celt from Canberra, Australia"
      1..>1041
        /gene="ND2"
      1..>1041

QY 3 gcaggcagtgagggtgattaaacacacacagctacgcaaaatcttagcatactctca 62
Db 490 GCAGGCAGTGTGAGTGAATAACCAACCCAGCTACGCAAAATCTTAGCATACTCTCA 549

QY 63 attaccacatagatgaataacacacacagctctaccgtacacccc 106
Db 550 ATTACCCACATAGGATGAATAATAGCAGTCTACCGTACACCC 593

RESULT 14
AF014888
LOCUS AF014888 1041 bp DNA PRI 06-MAY-1999
DEFINITION Homo sapiens NADH dehydrogenase subunit 2 (ND2) gene, mitochondrial
ACCESSION AF014888
VERSION AF014888.1 GI:3287315
KEYWORDS human.
SOURCE Mitochondrion Homo sapiens
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 1041)
AUTHORS Wise,C.A., Sraml,M. and Easteal,S.
TITLE Departure from Neutrality at the Mitochondrial NADH dehydrogenase
subunit 2 gene in Humans, but not in Chimpanzees
JOURNAL Genetics (1997) In press
REFERENCE 2 (bases 1 to 1041)
AUTHORS Wise,C.A.
TITLE Direct Submission
JOURNAL Submitted (21-JUL-1997) Human Genetics Group, John Curtin School of
Medical Research, The Australian National University, Mills Rd,
Canberra, ACT 0200, Australia
COMMENT On Jul 2, 1998 this sequence version replaced gi:2582068.
FEATURES
  source
    Location/Qualifiers
      1..1041
        /organism="Homo sapiens"
        /organelle="mitochondrion"
        /strain="Eur2"
        /db_xref="taxon:9606"
        /note="European: Anglo Celt from Canberra, Australia"
      1..>1041
        /gene="ND2"
      1..>1041
        /codon_start=1
        /transl_table=2
        /product="NADH dehydrogenase subunit 2"
        /protein_id="AAC25447.1"
        /db_xref="GI:3287316"
        /translation="MNPQAQVYISTIFAGTLTALSSHWFTWVGLMNMIAFIPVL
        TKMNPSTEAARKYFLQTATSMILLMAILFNMLSGOWTNTTNOYSSLMNMAM
        AMKLGAPFHFVPEVTQGTPLTSGLLLTWQKLAPISIMYQISPSLVSLTLTSL
        SIMAGSWGCGNOTQRLKILAYSSITHMGWMAVLDPYNNPNTILNLYIILTTAFLL
        LNLNSTTLLSRTWKNLTWLTPLIPSLTSLGGLPLTGLPKWAIIEETKKNLSL
        IIPTIMATILNLNLYFLRLIYSTISITLLPMSNVKMKWQFEHTKPTPLLIALT
        LLLPISPFMLMIT"
      325 a 349 c 100 g 267 t
      ORIGIN

Query Match 7.7%; Score 104; DB 40; Length 1041;
Best Local Similarity 100.0%; Pred. No. 9e-14;
Matches 104; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 gcaggcagtgagggtgattaaacacacacagctacgcaaaatcttagcatactctca 62
Db 490 GCAGGCAGTGTGAGTGAATAACCAACCCAGCTACGCAAAATCTTAGCATACTCTCA 549

QY 63 attaccacatagatgaataacacacacagctctaccgtacacccc 106
Db 550 ATTACCCACATAGGATGAATAATAGCAGTCTACCGTACACCC 593
```

SIMAGSWGGLNQTQLRKILAYSSITHMGMMAYLPYNPNMTILNLIIYIILTTAFLL
LNLSSTTLLLSRTNKNLTLWLPSTLLSLGGLPLPTGFLPKWAIIEEFTKNNSL
IIPITATITLLNLVFLRLIYSTITLLPMSNNVKMKQFHTKPTPELPTLIALT
LLLPISPFMLMIL" 324 a 349 c 101 g 267 t

Query Match 7.7% Score 104; DB 40; Length 1041;
Best Local Similarity 100.0%; Pred. No. 9e-14;
Matches 104; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 3 gcaggcagtgagggtgattaaacacacccagctacgcaaaatcttagcatactctcca 62
|||||
Db 490 GCAGGCAGTTGAGTGGATTAAACCAACCCAGCTACGCAAAATCTTAGCATACTCTCTCA 549
|||||

Qy 63 attaccacataggtgaataatagcagttctaccgtacacacc 106
|||||
Db 550 ATTACCCACATAGGATGAATAATAGCAGTTCTACCGTACACCC 593
|||||

RESULT 15
AF014889 1041 bp DNA PRI 06-MAY-1999
LOCUS Homo sapiens NADH dehydrogenase subunit 2 (ND2) gene, mitochondrial
DEFINITION gene encoding mitochondrial protein, partial cds.
ACCESSION AF014889
VERSION AF014889.1 GI:3287317
KEYWORDS human.
SOURCE Mitochondrion Homo sapiens
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 1041)
Wise, C.A., Sraml, M. and Easteal, S.
Departure from Neutrality at the Mitochondrial NADH dehydrogenase
subunit 2 gene in Humans, but not in Chimpanzees
Genetics (1997) In press
2 (bases 1 to 1041)
Wise, C.A.
AUTHORS Direct Submission
TITLE Submitted (21-JUL-1997) Human Genetics Group, John Curtin School of
Medical Research, The Australian National University, Mills Rd,
Canberra, ACT 0200, Australia
COMMENT On Jul 2, 1998 this sequence version replaced gi:2582070.
FEATURES
Location/Qualifiers
1..1041
/organism="Homo sapiens"
/organelle="mitochondrion"
/strain="Eur3"
/db_xref="taxon:9606"
/note="European: Anglo Celt from Canberra, Australia"
1..>1041
/gene="ND2"
1..>1041
/gene="ND2"
/codon_start=1
/transl_table=2
/product="NADH dehydrogenase subunit 2"
/protein_id="AAC25448.1"
/db_xref="GI:3287318"

gene
CDS
BASE COUNT 324 a 349 c 101 g 267 t
ORIGIN

TKKMPRTEAATKYLQATSMILLMAILFNMLSCGWTSTNQYSSLMIMAM
AMKLGAPFHFVPEVTQGTPLTGLLLTWQKAPISIMYQISPLNVSLLTSL
SIMAGSWGGLNQTQLRKILAYSSITHMGMMAYLPYNPNMTILNLIIYIILTTAFLL
LNLSSTTLLLSRTNKNLTLWLPSTLLSLGGLPLPTGFLPKWAIIEEFTKNNSL
IIPITATITLLNLVFLRLIYSTITLLPMSNNVKMKQFHTKPTPELPTLIALT
LLLPISPFMLMIL"

Query Match 7.7% Score 104; DB 40; Length 1041;

Best Local Similarity 100.0%; Pred. No. 9e-14;
Matches 104; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 3 gcaggcagtgagggtgattaaacacacccagctacgcaaaatcttagcatactctcca 62
|||||
Db 490 GCAGGCAGTTGAGTGGATTAAACCAACCCAGCTACGCAAAATCTTAGCATACTCTCTCA 549
|||||

Qy 63 attaccacataggtgaataatagcagttctaccgtacacacc 106
|||||
Db 550 ATTACCCACATAGGATGAATAATAGCAGTTCTACCGTACACCC 593
|||||

Search completed: May 15, 2000, 02:00:07
Job time: 11526 sec

" "

DB 204 CTGGTTCCAGCCCTCCAGAAAGTGAAGGCGTGGCTGAGCTCTTGAAGACTTATGAC 263
QY 427 aagacacgttcagaaagtctgctgaagtttgggtcagcagcctgcagggggagagctgc 486
DB 264 AGGACGACCTTCAGAAAGTGTGAAACTTGTAGTCGGGGCCCTGCTATGGGAAAGACTGC 323
QY 487 cagacaggtgtgacagctctggttgagcagcaccctgcagggagagcagctgggtgccc 546
DB 324 AGAAGAGCTGTGGAGCAACTTGGTCCAGCCCAACCTGTGAGAGAGCGTCTGGCCCTC 383
QY 547 ctgctgagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 606
DB 384 CTGCTGGCGGGACACACACCTCTGCTCAGCAGAGCTCTCGGCTGCCCTGCTAGTCTA 443
QY 607 aagcctgacacacttcagggaccagctccagggagcagcagcagcagcagcagcagcag 666
DB 444 AAGCCAGATGCTTCAGGAAGAGCTCCAGGAACCTGGCATTTCCCTCAGGATCTTAATTGA 503
QY 667 gacttgacagcagctggtatttgagcagcagcagcagcagcagcagcagcagcagcag 726
DB 504 GATTGGCCAGTTGGCAATTTGGAGTCAACGCCCTCTCTCAGACTCTGTAGCCCAACAG 563
QY 727 cagggggcctggctgcagctgtgctgacttctcgtgctgctgctgctgctgctgctgct 786
DB 564 CAGGGATCTCGTCTCAGCTGTCTTACTTCCGGTGGCGGTGAGCTGGCCATCTCA 623
QY 787 accagtgccctggctgcctgcagcagcagcagcagcagcagcagcagcagcagcagcag 846
DB 624 ACCAGCGCTCAGTCCCGCTCCCTGCAACCGAGTGTCTCATCAGCTGAAGCTCACAGAT 683
QY 847 ggtcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 906
DB 684 GATCTGCACACCCCTTCAGAGTGGCCATAGCCAAATTTTCAGAGCTGGCGGTACAGTGA 743
QY 907 gccctggtcttaagagagagagcagcagcagcagcagcagcagcagcagcagcagcag 966
DB 744 GCCTTGTCTTTAAGGAGATGGCAGAACTGGAGAAAGTGTGAGCGCAAACTGACGAG 803
QY 967 tgaccctcaactgaccagt 986
DB 804 TGACTGAACCCCTGGTACTGT 823

RESULT 2

V18891
ID V18891 standard; cDNA; 701 BP.
AC V18891;
DT 09-JUN-1998 (first entry)
DE Human Hypertension related calcium regulated gene (HcARG) cDNA.
KW Hypertension related calcium regulated gene; HcARG; human parathyroid;
KW extracellular calcium concentration; antibody; hypertension; ss;
KW hyperthyroidism; osteoporosis; heart failure; diabetes; stroke;
KW cancer; inflammatory disease; asthma.
OS Homo sapiens.
FT Key Location/Qualifiers
FT misc_feature 224..245
FT /tag= a
FT /note= "EF-hand like motif"
FN W09749807-A2.
PD 31-DEC-1997.
PF 23-JUN-1997; CA0439.
PR 21-JUN-1996; US-667495.
PA (GOSS)/ GOSSARD F.
PA (HAME)/ HAME P.
PA (LEWA)/ LEWANCZUK R.
PA (TREM)/ TREMBLAY J.
PI Gossard F, Hame P, Lewanczuk R, Tremblay J;
DR WPI; 98-0771/1/07.
PT Hypertension related calcium regulated gene - useful to develop
PT products to treat or detect, e.g. hypertension, stroke,
PT osteoporosis, heart failure, cancer, diabetes or asthma
PS Claim 3; Pages 27-28; 46pp; English.

CC This nucleic acid sequence encodes the hypertension related calcium
CC regulated gene (HcARG), which was isolated from the human parathyroid.
CC Its expression is regulated by extracellular calcium concentration.
CC The sequence displays 75 percent homology to the rat form, with
CC conserved areas such as the EF-hand like motifs and the initiating
CC codon. An antibody against the protein, can be used to detect or
CC modulate (e.g. enhance or inhibit) abnormal calcium levels. They can
CC specifically be used to detect or treat, e.g. hypertension,
CC hyperthyroidism, osteoporosis, heart failure, diabetes, cancer,
CC inflammatory disease, and asthma.
CC Sequence 701 BP; 137 A; 204 C; 217 G; 139 T;
SQ

Query Match 25.0%; Score 338.8; DB 1: Length 701;
Best Local Similarity 72.2%; Pred. No. 8e-80;
Matches 659; Conservative 0; Mismatches 30; Indels 224; Gaps 8;

QY 100 acaaccggagggcagagaccgtccacaggggacagggagcggcgtcgctccgctccg 159
DB 8 AAAACCTGTGAGGCGCAGACCGTCCACAGAGACAGGGAGCGGCTCGCTCCGCTCCG 67
QY 160 cgtgatacctcaaggtctcgtgggttccccgcgcgcaccccgacccgacccgacgaagccag 219
DB 68 -GCTGATCTTCAGGCTCTCGGCTTTCCCNCCGCCACCCCGGACGCCGACGAAGCCAG 126
QY 220 cgagctcctcagcctcagggcagctgcatctggaacagcctcctgggctggctgacaaa 279
DB 127 CGAGCTCTCAGTCT-----CGACCTCTGGGCTGGCTGATCAAA 166
QY 280 gagaaacagcagcaatgtctgtgggggtgcaactccatacctgcatcctcctggt 339
DB 167 GAGGAACGACGAGCAATGTCTGCTTTGGGACACTGCAACTCCATACCTGCAATCCTGGT 236
QY 340 gatagtcacagtggccagtgagttcttggggggccagcttccctccagaggtggcagca 399
DB 227 GATAGTCACAGTGGCCGAGTGAGTTCTTGGGGGCCAGCTTCTCCAGAGGTGGCAGCA 286
QY 400 atggccgggtactataggggacctagacagcagcagctcagaaagttgc-tgaagttgt 458
DB 287 ATGGCCCAAGCTACTAGGGGACCTAGACAGAGACAGCTTCAGAAAGTTGCTTGAAGTTGT 346
QY 459 ggtcagcagcctgcagggggagggactgcagagcaggtgtgcagcgtcttggggtcagcgc 518
DB 347 GGTACAGACCTGCAGGGGGAGGATTGCCGAGAGTGTTCGAGCGTNTTTGGGGTCAGCGC 406
QY 519 caactgcgggagagcagctgggtgcccctgctggcagggcagcagcagcagcagcagcagca 578
DB 407 CAACCTGCGGAGGAGCAGTTNGGTTGCCT----- 437
QY 579 ggcctcctgctgccccccaccagcctgaagcctgcacaccttcaggaggaccagctccagga 638
DB 437 ----- 437
QY 639 gctctgcatcccccaagacctggtcggggagctggccagcgtgtatttgggagccagcg 698
DB 437 ----- 437
QY 699 gccctccttgattctgtggccccagcagcagggggcctggctgcgcagctgttgcgactt 758
DB 437 -----TGTTGGCAGGTT 448
QY 759 tcggtggcgggtggatgtag-caatctccacc-agtgccttggctgcctccctgcagccg 816
DB 449 TCGGTGGCGGGTGGATGTAGCAAACTCCACCAGTGGCTGGCTCGCTCCCTCGCAGCGC 508
QY 817 agcgtcctgatgacagctg-aagcttccagatgggtcagcata-cgcgtttgaggtcccca 874
DB 509 AGCGTCTCTGATGACGCTGTAGCTTTTCAGTGGGTGAGCATAACCGCTTTGAGGTCCCA 568
QY 875 cagcaagcttcagagcagctgcgttacagcgtggcctcctggtcctaaagagagatggcagatc 934
DB 569 CAGCCCAAGTTCCAGAGCTGCCGTACAGCGTGGCCCTGGTCTCTAAAGGAGATGGCAGATC 628

QY 935 tggagaagaggtgtgagcgcagactgcagagactgacccctcacttgaccagctcccatcca 994
 Db 629 TGGAGAAGAGGTGTGAGCGCAGACTGACCCCTCACTTGACGAGTCCCATCCA 688
 QY 995 gatccggttggga 1007
 Db 689 GATCCGGCTTGGG 701

RESULT 3

T25721
 ID T25721 standard; cDNA to mRNA; 186 BP.
 AC T25721;
 DT 10-OCT-1996 (first entry)
 DE Human gene signature HUMGS07932.
 KW Gene signature; messenger RNA; mRNA; relative abundance; frequency;
 KW human; cloning; mapping; non-biased library; diagnosis; detection;
 KW cell typing; abnormal cell function; ss.
 OS Homo sapiens.
 PN W09514772-AL.
 PD 01-JUN-1995.
 PF 11-NOV-1994; J01916.
 PR 12-NOV-1993; JP-355504.
 PA (MATS/) MATSUBARA K.
 PA (OKUB/) OKUBO K.
 PI Matsubara K, Okubo K;
 DR WPI; 95-206931/27.
 PT Identifying gene signatures in 3'-directed human cDNA library - e.g.
 PT for diagnosis of abnormal cell function, by preparing cDNA that
 PT reflects relative abundance of corresp. mRNA in specific human
 PT tissues
 PS Claim 1; Page 1914; 2245pp; Japanese.
 CC A single-stranded DNA (or its complementary strand or the corresp.
 CC double-stranded DNA) which comprises one of the 7837 "GS" sequences
 CC given in T19001-T26837 and which is able to hybridise to part of
 CC human genomic DNA, cDNA or mRNA is claimed. The GS (Gene Signature)
 CC sequences were obtained from 3'-directed cDNA libraries prepared
 CC from various human tissues; synthesis of cDNA was initiated from the
 CC 3'-end of mRNA by using poly(T) as the sole primer. Since the 3'-
 CC untranslated sequence is unique to a particular mRNA species, almost
 CC all the 3'-oriented cDNAs hybridise with specific mRNAs. Each library
 CC is constructed so as to reflect accurately the relative abundance of
 CC different mRNAs in the particular tissue from which it was derived.
 CC The appearance frequency of a given GS in a cDNA library can be
 CC determined (esp. using primers and probes derived from the GS
 CC sequences) as a means of diagnosing abnormal cell function or for
 CC recognising different cell types.
 SQ Sequence 186 BP; 46 A; 40 C; 43 G; 54 T;

Query Match 13.0%; Score 175.8; DB 1; Length 186;
 Best Local Similarity 97.8%; Pred. No. 3e-37;
 Matches 177; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 995 gatccggttggcagcagcactgaagatgggccaaagtgcagctgactctccacgaca 1054
 Db 1 GATCCGCTTGGACAGCACCTGAGATGGTGCCAAAGTCAGCTGACTCTTCCACGACA 60
 QY 1055 gccctggcctcccatcagcagcagctcttcagtgagtggttgacgtaattatgtatt 1114
 Db 61 GCCTGCCCCCTCCCATGAGCAGCGCTCTTCAGTGAGTGTGACGTAATTATGTAGTTT 120
 QY 1115 tcgtttaattgaaaagagagctatgccttttttttttttttttttttttttttttttt 1174
 Db 121 TCTGTTAATTGAAAAGNAGACTATGCCTTTTCTTTTCTTTTCTTTTCTTTTCTTTT 180
 QY 1175 a 1175
 Db 181 A 181

RESULT 4

V86703

ID V86703 standard; cDNA; 448 BP.
 AC V86703;
 DT 27-APR-1999 (first entry)
 DE EST clone BG461.
 KW Expressed sequence tag; secreted protein; haematopoiesis regulator;
 KW tissue growth; activin; inhibin; tumour invasion suppressor; EST; human;
 KW chemotaxis; chemokinesis; haemostasis; gene therapy; thrombolysis;
 KW receptor; ligand; anti-inflammatory; tumour inhibitor; ds.
 OS Homo sapiens.
 PN W09845435-A2.
 PD 15-OCT-1998.
 PF 10-APR-1998; U06954.
 PR 10-APR-1997; US-835913.
 PA (GEMY) GENETICS INST INC.
 PI Agostino MJ, Jacobs K, Lavallie ER, McCoy JM, Merberg D,
 PI Racie LA, Spaulding V, Treacy M;
 DR WPI; 99-070076/06.
 PT New polynucleotides encoding human secreted proteins - derived from
 PT e.g. human blood, kidney, foetal lung, placenta, testes, brain,
 PT ovary, pituitary, retina and colon cDNA libraries
 PS Claim 1; Page 321; 633pp; English.
 CC This sequence represents an expressed sequence tag (EST), and is a
 CC polynucleotide of the invention. The polynucleotides of the invention are
 CC all secreted EST sequences isolated from a variety of human tissue
 CC sources. The EST sequences and proteins encoded by them are predicted to
 CC have useful biological activities which would make them suitable for
 CC treating, preventing or ameliorating medical conditions in humans and
 CC animals, although no supporting data is given. Suggested activities
 CC include nutritional activity, immune stimulating or suppressing activity,
 CC haematopoiesis regulating activity, tissue growth activity,
 CC activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
 CC and thrombolytic activity, receptor/ligand activity, anti-inflammatory
 CC activity, cadherin/tumour invasion suppressor activity, tumour inhibition
 CC activity. The EST sequences are also stated to be useful for gene
 CC therapy.
 SQ Sequence 448 BP; 144 A; 137 C; 52 G; 115 T;

Query Match 7.7%; Score 104.4; DB 1; Length 448;
 Best Local Similarity 99.1%; Pred. No. 2.4e-18;
 Matches 105; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 3 gcaggcagtggtgattgaataacacacccagctacgcacaaattcttagcatactctca 62
 Db 252 GCAGGCGAGTGTGAGTGGATTAAACCAACCCAGCTACGCAAAATCTAGTACTCTCA 311
 QY 63 attaccacacatgagtgatgaataatagcagttctacgtacacaccccg 108
 Db 312 ATTATCCACATAGGATGAATATAGCAGTTCCTACGTACACCCCTG 357

RESULT 5

V38226
 ID V38226 standard; DNA; 350 BP.
 AC V38226;
 DT 28-SEP-1998 (first entry)

DE Human mucosal addressin cell adhesion molecule-1 gene exon 3.
 KW Mucosal addressin cell adhesion molecule-1; MACAM-1; human;
 KW cancer; tumour; inflammation; transplant rejection; arthritis;
 KW rheumatoid arthritis; infection; inflammatory bowel disease;
 KW autoimmune disease; experimental autoimmune encephalitis;
 KW dermatosis; diagnosis; therapy; ds.
 OS Homo sapiens.
 PN W09820110-A1.
 PD 14-MAY-1998.

PF 01-NOV-1996; U17549.
 PR 01-NOV-1996; WO-U17549.
 PA (HUMA-) HUMAN GENOME SCI INC.
 PA (UYAU-) UNIV AUCLAND.
 PI Greene JM, Krissansen GW, Leung EYF, Ni J, Ruben SM;
 DR WPI; 98-286926/25.

PT Addressin cell adhesion molecules - used to develop products for
 PT detection of inflammatory conditions or cancer and for treating or

CC and/or cellular response, specifically of Th1 type, particularly
 CC including induction of interleukin-12 (IL-12) production. They may thus
 CC be used more generally to treat any condition (e.g. bacterial, viral or
 CC protozoal infection, or cancer) which responds to IL-12.
 SQ Sequence 1771 BP; 279 A; 552 C; 672 G; 268 T;

Query Match 3.2%; Score 44; DB 1; Length 1771;
 Best Local Similarity 45.8%; Pred. No. 0.031;
 Matches 152; Conservative 0; Mismatches 180; Indels 0; Gaps 0;

QY 376 cagcttcctcagagtggtgcagcaatggccggctactagggagcctagacagagcagc 435
 DB 1050 CGCGCCCTCGATCAGTCGCGCTTTCACCTCCCTGCTAGTCGACGTCGCGGCGG 991
 QY 436 ttcagaaagtgcgaagtgtgtcagcagctgcagggggagagcagtcgcgagacggt 495
 DB 990 CTCATCTTTTCCCTCCTCCTCAGCTCCAGTCCGCGCAGACGCGCTCCTCTCGAG 931
 QY 496 gtgcagcgtcttgggtgcagcgaacctgcggaggagcagctgggtgcccctgtggca 555
 DB 930 GGCCTGGAGCGGCGCTCTCGGCTCTGTCATGCGCTCGAGGGCGGCGCTCTCTCGAG 871
 QY 556 ggcagtcacacactgctccagcagcgcctccgtctgccccccacacagcctgaagcctgac 615
 DB 870 GGCCTGGACACGGGCGCTCTCGGCTCTGTCATGACATCCAGGGGCGGCGCTCTCTCGGC 811
 QY 616 accttcagggaccagctccagagcctctgcctcccccagacccctggtggagacttgccc 675
 DB 810 CGCCTCCAGCTCCGCGGAGACGCGCTCCTCTCGAGGGCGGCGGCGGCGCTCTCTGTC 751
 QY 676 agcgtggtattggagcagcagcggccctccct 707
 DB 750 GGCCTCGTGCATGACATCCAGCGGCGGCGCT 719

RESULT 8
 V38201
 ID V38201 standard; DNA; 1536 BP.
 AC V38201;
 UJ 28-SEP-1998 (first entry)
 DE Human mucosal addressin cell adhesion molecule-1(a) DNA.
 KW Mucosal addressin cell adhesion molecule-1; MadCAM-1(a); human;
 KW cancer; tumour; inflammation; transplant rejection; arthritis;
 KW rheumatoid arthritis; infection; inflammatory bowel disease;
 KW autoimmune disease; experimental autoimmune encephalitis;
 KW dermatosis; diagnosis; therapy; ds.
 OS Homo sapiens.
 FH Key Location/Qualifiers
 FT CDS 1..1149
 FT sig_peptide /*tag= a
 FT 1..49
 FT mat_peptide /*tag= b
 FT 52..1146
 FT polyA_signal /*tag= c
 FT 1507..1512
 FT /*tag= d
 PN WO9820110-A1.
 PD 14-MAY-1998.
 PF 01-NOV-1996; WI7549.
 PR 01-NOV-1996; WO-U17549.
 PA (HUMA-) HUMAN GENOME SCI INC.
 PA (YIAU-) UNIV AUCKLAND.
 PI Greene JM, Krissansen GW, Leung EYF, Ni J, Ruben SM;
 DR WPI; 98-286926/25.
 DR P-PSDB: W60607.
 PT Addressin cell adhesion molecules - used to develop products for
 PT detection of inflammatory conditions or cancer and for treating or
 PT preventing inflammatory conditions
 PS Claim 2; Page 88-90; 164pp; English.
 CC This nucleic acid molecule, designated clone HEBC23 and deposited
 CC as ATCC 97759, codes for human mucosal vascular addressin cell
 CC adhesion molecule MadCAM-1(a) (see W60607), a novel cell surface

CC adhesion molecule that shows homology to murine MadCAM-1. The
 CC invention relates to human MadCAM-1(a) as well as 4 splice variants,
 CC designated MadCAM-1(b), -1(c), -1(d) and -1(e) (see V38201-05).
 CC These 5 nucleic acid molecules were discovered in a cDNA library
 CC derived from human fetal brain cells following a database search
 CC for human ESTs having homology for mouse MadCAM-1. The genes were
 CC also identified in cDNA libraries from the small intestine, colon,
 CC spleen and pancreas. The invention also provides MadCAM-1(a-e)
 CC proteins (see W60607-11) vectors, host cells, recombinant methods
 CC of producing the polypeptides, as well as methods for identifying
 CC agonists and antagonists of activity, diagnostic methods for
 CC detecting cancer or a pathological inflammatory condition, and
 CC therapeutic methods for treating an individual in need of a
 CC reduction in activity of any of MadCAM-1(a-e). Genomic DNA
 CC molecules comprising the 5' untranslated region and exons 1-5 (see
 CC V38223-28) which, in various combinations, comprise the coding
 CC region of any of the MadCAM-1 splice variants are also claimed.
 CC The novel human MadCAM-1 polypeptides can be used as a target for
 CC the diagnosis and treatment of inflammation conditions such as
 CC transplant rejection, arthritis, rheumatoid arthritis, infection,
 CC dermatosis, inflammatory bowel disease, and autoimmune disease,
 CC including chronic relapsing experimental autoimmune encephalitis.
 SQ Sequence 1536 BP; 263 A; 571 C; 442 G; 260 T;

Query Match 3.1%; Score 41.8; DB 1; Length 1536;
 Best Local Similarity 46.4%; Pred. No. 0.11;
 Matches 136; Conservative 0; Mismatches 157; Indels 0; Gaps 0;

QY 372 gggccagcttcctccagagtggtgcagcaatggccggctactagggagcctagacagag 431
 DB 375 GGTCCCTGGTGACCCGAGGAGTGCCCTGTACGGCCACAAAGTCACGCCCTGGACCCCAA 434
 QY 432 cagcttcagaaagtgtgctgaagttgtgtcagcagcctgcggaggagcagctgggtgctcct 491
 DB 435 CGCGCTCTCCTCTCCCTGCTCTCGGGGCGCAGGAATCTGGAGGGGCGCAAGCCCTGGG 494
 QY 492 cgggtgcagcgtcttgggtgcagcctgcggaggagcagcagctgggtgctcctgct 551
 DB 495 CCGGAGGTGACAGGAGGAGGAGGAGGCCCCAGGGGGGACGAGCGTGTCTTTCAGGGT 554
 QY 552 ggcagggcatgcacacactgctccagcagcgtccctcctgctccccccaccagcctgaagcc 611
 DB 555 GACAGAGCGGTGGCGGCTGCGCCCTGGGAGCCCTGTCCCGCCGCCCTCTACTGCCCA 614
 QY 612 tgacaccttcagggagcagctccagagctgtgcatcccccaagacctggtcg 664
 DB 615 GGCCACGATGAGGCTGCCTGGCTTGGAGCTCAGCCACCGCCAGGCGCATCCCG 667

RESULT 9
 V38202
 ID V38202 standard; DNA; 1488 BP.
 AC V38202;
 DT 28-SEP-1998 (first entry)
 DE Human mucosal addressin cell adhesion molecule-1(b) DNA.
 KW Mucosal addressin cell adhesion molecule-1; MadCAM-1(b); human;
 KW cancer; tumour; inflammation; transplant rejection; arthritis;
 KW rheumatoid arthritis; infection; inflammatory bowel disease;
 KW autoimmune disease; experimental autoimmune encephalitis;
 KW dermatosis; diagnosis; therapy; ds.
 OS Homo sapiens.
 FH Key Location/Qualifiers
 FT CDS 1..1101
 FT sig_peptide /*tag= a
 FT 1..49
 FT mat_peptide /*tag= b
 FT 52..1098
 FT polyA_signal /*tag= c
 FT 1459..1464
 FT /*tag= d
 PN WO9820110-A1.
 PD 14-MAY-1998.

cancer; anorexia; bulimia; Parkinson's disease; cardiovascular disease; neurological disorder; Huntington's disease; gene therapy; gene mapping; Gilles de la Tourette's syndrome; ss.

OS Homo sapiens.
PN EP-894863-AL.
PP 03-FEB-1999.
PP 21-JUL-1998; 305794.
PP 20-FEB-1998; US-027064.
PP 28-JUL-1997; US-053924.
PR (SMK) SMITHKLINE BEECHAM CORP.
PA Bergsma DJ, Shabon U;
PI WPI; 99-108353/10.
DR P-PSDB; W94998.
DR New serine-threonine kinase (HTLAR33) polypeptides and
PPT polynucleotides - useful as diagnostic reagents and for prevention
PPT and treatment of bone loss, neurological and inflammatory disorders
PPT and cancer, HIV infections and angina pectoris
PS Claim 13; Page 15-17; 30pp; VAK-1.
CC The invention relates to a VAK-1 related serine-threonine protein kinase
CC polypeptide (HTLAR33). Host cells transformed with an expression vector
CC comprising the HTLAR33 nucleic acid are used for the recombinant
CC production of the protein. HTLAR33 polynucleotides and polypeptides are
CC useful for are useful for diagnosing susceptibility to diseases and for
CC screening for antagonists, agonists. These can be used in treatment to
CC enhance (agonist) or block (antagonist) HTLAR33 activity. Diseases
CC diagnosed prevented or treated include: bone loss and inflammatory
CC disorders including osteoporosis, Adult Respiratory Disease Syndrome
CC (ARDS), Rheumatoid arthritis, osteoarthritis, Inflammatory Bowel Disease
CC (IBD), psoriasis, dermatitis, asthma, allergies; bacterial, fungal,
CC protozoan and viral infections, especially those caused by HIV-1 or
CC HIV-2; HIV-associated cachexia and other immunodeficient disorders;
CC septic shock; pain; injury; cancers; anorexia; bulimia; Parkinson's
CC disease; cardiovascular disease including restenosis, atherosclerosis,
CC acute heart failure, myocardial infarction; hypotension; hypertension;
CC urinary retention; angina pectoris; ulcers; benign prostatic hypertrophy;
CC and psychotic and neurological disorders, including schizophrenia, manic
CC depression, anxiety, delirium, dementia, severe mental retardation and
CC dyskinesias, such as Huntington's disease or Gilles de la Tourette's
CC syndrome. HTLAR33 polypeptides are also useful for mapping genes to
CC chromosomes, allowing gene inheritance to be studied through linkage
CC analysis.
CC Sequence 1338 BP; 238 A; 495 C; 373 G; 232 T;
SQ

Query Match 3.0%; Score 40.2; DB 1; Length 1338;
Best Local Similarity 53.5%; Pred. No. 0.28;
Matches 84; Conservative 0; Mismatches 73; Indels 0; Gaps 0

QY 526 ccggaggagcagctgggtccctgtcgtgcaggatgcacacatgctccagcaggccctc 585
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 719 CCGGAGCTGATGATGTGACCTGTCGCTGGCGGCCCTGCTGACTGCTCCACCTCACCAG 778
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 586 cgtctgccccccaccagcctgaagcctgcacacttcaggaccagctccaggagctctgc 645
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 779 CGCCTGCCCCCAGCACCCCGCTGCTCAGCCCTCCGGACTCGGATGACTGGAGTGCTC 838
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 646 atcccccaagacbtgctgggggacttggccagcgtgg 582
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 839 CACCCTCCCGCCTCCTGATGATGACCTTGCCACTCTGGG 875
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 15
X17774
ID X17774 standard; cDNA; 2394 BP.
AC X17774;
DC 12-MAY-1999 (first entry)
DE VAK-1 related serine/threonine protein kinase-HTLAR33 encoding cDNA.
KW YAK-1: serine-threonine protein kinase; HTLAR33; bone loss; ARDS;
KW inflammatory disorder; osteoporosis; Adult Respiratory Disease Syndrome;
KW arthritis; psoriasis; dermatitis; asthma; allergy; infection; HIV-1;
KW HIV-2; cachexia; immunodeficient disorder; septic shock; pain; injury;
KW cancer; anorexia; bulimia; Parkinson's disease; cardiovascular disease;
KW neurological disorder; Huntington's disease; gene therapy; gene mapping;
KW

W09725351-A2.
 17-JUL-1997. PF
 03-JAN-1997; U00291.
 04-JAN-1996; US-582740. PR
 (LEUK-) LEUKOSITE INC. PA
 Schwender CF, Shroff HN;
 WPI: 97-384978/35.
 P-P5DB; Y02068.
 New polypeptide(s) for inhibiting MadCAM-1-mediated interactions -
 useful for treatment of diseases associated with leukocyte
 infiltration of tissues, especially inflammatory bowel disease and
 insulin-dependent diabetes mellitus
 Disclosure; Fig 4: 108pp; English.
 The specification describes inhibitors of mucosal addressing cell
 adhesion molecule-1 (MADCAM-1) mediated interactions. These inhibitors
 comprise peptides which mimic the conserved amino acid motif LDTSL of
 MADCAM-1, and which have groups bonded to the N and C terminals
 The inhibitors can be used for the treatment of a disease associated
 with leukocyte infiltration of tissues expressing the MADCAM-1,
 especially inflammatory bowel disease and insulin-dependent diabetes
 mellitus. Inflammatory bowel diseases which can be treated include
 ulcerative colitis, Crohn's disease, ileitis, coeliac disease,
 nontropical sprue, enteropathy associated with seronegative
 arthropathies, microscopic or collagenous colitis, eosinophilic
 gastroenteritis, or pouchitis resulting after proctocolectomy and
 ileoanal anastomosis. The inhibitors can also be used to inhibit the
 bonding of a cell expressing a ligand for MADCAM-1 on the cell surface to
 MADCAM-1. They are also useful in diagnostic and research applications,
 e.g. as immunogens (when conjugated to a suitable carrier) to induce the
 formation of antibodies which selectively bind MADCAM-1. These antibodies
 can in turn be used to identify cells expressing MADCAM-1 on their cell
 surface or to detect MADCAM-1 in a sample. The compounds can also be
 labelled and used to detect alpha4-beta-integrin and/or to quantitate
 expression of this the surface of cells. The present sequence
 encodes human MadCAM-1.

	295 A;	610 C;	454 G;	265 T;
	Sequence	1624 BP;		

Query Match	3.1%;	Score 41.6;	DB 1;	Length 1624;
Best Local Similarity	46.8%;	Pred. No. 0.13;		
Matches 131;	Conservative	O: Mismatches 149;	Indels 0;	Gaps 0;

```

Qy   385 ccaagggtggcagcaatgcccggtactatggggacccctagacaggagcacgttcagaag 444
      ||| ||||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| ||| |||||
Db   388 CGGAGGGTGGCCGTATAGGCCCAAAAGTCACGCCCGGTGGACCCCAACGGCGTCTCCTTC 447
      ||| ||||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| ||| |||||
Qy   445 ttgcgaagtgttggtcagcagcgtccaggggggggacctgcgcgagacggtgtgcagcgt 504
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db   448 TCCTCTGCTGCTCGGGGGCCAGGAACGTGAGGGGGGGGCACAGCCCTGGGCCGAGGTGCAG 507
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy   505 ctctggggtcagcccaactgcgcgagagcagctgggtgccttgcctggcaggcatgcac 564
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db   508 GAGGAGGAGGAGAGCCCCAGGGGNCAGACGCTGCTTTTCAGGGTGACAGAGCGCTGG 567
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy   565 acactgtctccagcagccctccgctgtgcccccccaccagcctgaagcctgacaccttaagg 524
      |||| | |||| | |||| | |||| | |||| | |||| | |||| | |||| | |||| |
Db   568 CGGTCGCCGCCCTGGGGACCCTGTCCGCCGCCCTCTACTGCCAGGCCACGATGAGG 627
      |||| | |||| | |||| | |||| | |||| | |||| | |||| | |||| | |||| |
Qy   625 gaccagctccagagctgtgatcccccacaagacctggtcg 664
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db   628 CTCCTCGGCTTGAGCTCAGCCNCCCAGGCCATCCCG 667
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

```

RESULT 14
 X1775
 ID X1775 standard; cDNA; 1338 BP.
 AC X1775;
 DT 12-MAY-1999 (first entry)
 DE HTLAR33 encoding cDNA (EST derived sequence).
 KW YAK-1; serine-threonine protein kinase; HTLAR33; bone loss; ARDS;
 KW Inflammatory disorder; osteoporosis; Adult Respiratory Disease Syndrome;
 KW arthritis; psoriasis; dermatitis; asthma; allergy; infection; HIV-1;
 KW HTV-2; cachexia; immunodeficient disorder; septic shock; pain; injury;

KW Gilles de la Tourette's syndrome; ss.
OS Homo sapiens.
PN EP-894863-A1.
PD 03-FEB-1999.
PF 21-JUL-1998; 305794.
PR 20-FEB-1998; US-027064.
PR 28-JUL-1997; US-053924.
PA (SMIK) SMITHKLINE BEECHAM CORP.
PI Bergsma DJ, Shabon U;
DR WPI; 99-108353/10.
DR P-PSDB; W94997.
PT New serine-threonine kinase (HTLAR33) polypeptides and
PT polynucleotides useful as diagnostic reagents and for prevention
PT and treatment of bone loss, neurological and inflammatory disorders
PT and cancer, HIV infections and angina pectoris
PS Claim 2; Page 13-15; 30pp; English.
CC This cDNA encodes a YAK-1 related serine-threonine protein kinase
CC polypeptide (HTLAR33). Host cells transformed with an expression vector
CC comprising the HTLAR33 nucleic acid are used for the recombinant
CC production of the protein. HTLAR33 polynucleotides and polypeptides are
CC useful for are useful for diagnosing susceptibility to diseases and for
CC screening for antagonists, agonists. These can be used in treatment to
CC enhance (agonist) or block (antagonist) HTLAR33 activity. Diseases
CC diagnosed, prevented or treated include: bone loss and inflammatory
CC disorders including osteoporosis, Adult Respiratory Disease Syndrome
CC (ARDS), Rheumatoid arthritis, osteoarthritis, Inflammatory Bowel Disease
CC (IBD), psoriasis, dermatitis, asthma, allergies; bacterial, fungal,
CC protozoan and viral infections, especially those caused by HIV-1 or
CC HIV-2; HIV-associated cachexia and other immunodeficient disorders;
CC septic shock; pain; injury; cancers; anorexia; bulimia; Parkinson's
CC disease; cardiovascular disease including restenosis, atherosclerosis,
CC acute heart failure, myocardial infarction; hypotension; hypertension;
CC urinary retention; angina pectoris; ulcers; benign prostatic hypertrophy;
CC and psychotic and neurological disorders, including schizophrenia, manic
CC depression, anxiety; delirium, dementia, severe mental retardation and
CC dyskinesias, such as Huntington's disease or Gilles de la Tourette's
CC syndrome. HTLAR33 polypeptides are also useful for mapping genes to
CC chromosomes, allowing gene inheritance to be studied through linkage
CC analysis.
SQ Sequence 2394 BP; 478 A; 801 C; 676 G; 439 T;

Query Match 3.0%; Score 40.2; DB 1; Length 2394;
Best Local Similarity 53.5%; Pred. No. 0.35;
Matches 84; Conservative 0; Mismatches 73; Indels 0; Gaps 0;
QY 526 ccggaggagcagctgggtgcccctgtggtgaggatgcacacactgtccagcaggccctc 585
Db 1775 CCGGAGCTGATGGATGTGAGCTGTGTGGCGGCTGTGACTGCTCCCCACCTCACCAG 1834
QY 586 cgtctgccccccaccagcctgaagcctgacaccttcaggagaccagctccagagctctgc 645
Db 1835 CGCCTGCCCCCAGACCCGGCTGCTCAGCCCTCCGGACTCGGATGACTGGAGGTGTC 1894
QY 646 atcccccaagacctgctcgggagcttggccagcgtgg 682
Db 1895 CACCCCTCCGCTCTCTGATGACCTGCCCACCTCGG 1931

Search completed: May 15, 2000, 01:46:13
Job time: 9837 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 15, 2000, 01:50:08 ; Search time 76.12 Seconds
(without alignments)
2260.217 Million cell updates/sec

Title: US-09-223-796-3

Perfect score: 1355

Sequence: 1 gggcaggcagtgaggtaga.....gtgttcaggcaggccgg 1355

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 226296 seqs, 63486255 residues

Total number of hits satisfying chosen parameters: 452592

Minimum DB seq length: 0

Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database : Issued_Patents_NA:*

1: /cgn2_6/ptodata/2/ina/5A_COMB.seq:*

2: /cgn2_6/ptodata/2/ina/5B_COMB.seq:*

3: /cgn2_6/ptodata/2/ina/5C_COMB.seq:*

4: /cgn2_6/ptodata/2/ina/5D_COMB.seq:*

5: /cgn2_6/ptodata/2/ina/6_COMB.seq:*

6: /cgn2_6/ptodata/2/ina/6_COMB.seq:*

7: /cgn2_6/ptodata/2/ina/6_COMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	44	3.2	1771	3	US-08-533-669A-7
C 2	44	3.2	1771	4	US-08-511-872-1
C 3	39.6	2.9	1221	7	5212296-16
C 4	39.6	2.9	1879	7	5212296-5
C 5	38.6	2.8	5173	1	US-08-242-677-1
C 6	38.4	2.8	49272	2	US-08-614-770A-1
C 7	37.8	2.8	3489	4	US-08-728-323A-1
C 8	37.8	2.8	3207	3	US-08-770-379-20
C 9	37.4	2.8	3435	2	US-08-366-577-1
C 10	37.4	2.8	3435	6	PCT-US96-00005-1
C 11	35.8	2.6	864	1	US-08-396-650-2
C 12	35.8	2.6	864	2	US-08-788-626-2
C 13	35.8	2.6	921	1	US-08-396-650-3
C 14	35.8	2.6	921	2	US-08-396-650-4
C 15	35.8	2.6	921	1	US-08-768-626-3
C 16	35.8	2.6	921	2	US-08-768-626-4
C 17	35.8	2.6	1712	1	US-08-171-299B-1
C 18	35.6	2.6	432	2	US-08-642-255-48
C 19	35.6	2.6	756	2	US-08-642-255-50
C 20	35.6	2.6	1335	6	PCT-US91-06532-1
C 21	35.4	2.6	432	2	US-08-642-255-48
C 22	35.4	2.6	472	4	US-08-975-316-28
C 23	35.2	2.6	3234	1	US-08-264-534-31
C 24	35.2	2.6	3234	2	US-08-083-590A-10
C 25	35.2	2.6	3234	3	US-08-465-500-31
C 26	35.2	2.6	3234	3	US-08-346-128-31
C 27	34.6	2.6	756	2	US-08-642-255-50

28	34.6	2.6	2569	2	US-08-631-607-1	Sequence 1, Appl
29	34.6	2.6	3684	4	US-08-760-075A-17	Sequence 17, Appl
C 30	34.4	2.5	330	1	US-07-849-389-6	Sequence 6, Appl
C 31	34.4	2.5	2793	2	US-08-209-747-1	Sequence 1, Appl
C 32	34.4	2.5	2793	2	US-08-458-298-1	Sequence 1, Appl
C 33	34.4	2.5	12412	2	US-08-390-878-18	Sequence 18, Appl
C 34	34.2	2.5	895	5	US-08-650-766-4	Sequence 4, Appl
C 35	34.2	2.5	1171	5	US-08-650-766-5	Sequence 5, Appl
C 36	34.2	2.5	1678	5	US-08-650-766-2	Sequence 2, Appl
C 37	34.2	2.5	3318	5	US-08-650-766-3	Sequence 3, Appl
C 38	34.2	2.5	3385	5	US-08-650-766-1	Sequence 1, Appl
C 39	34.2	2.5	4257	3	US-08-690-473-1	Sequence 1, Appl
C 40	34.2	2.5	12001	2	US-08-458-568A-11	Sequence 11, Appl
C 41	34	2.5	1426	3	US-08-284-465-2	Sequence 2, Appl
C 42	34	2.5	1426	3	US-08-284-465-7	Sequence 7, Appl
C 43	33.3	2.5	43280	3	US-08-804-227C-1	Sequence 1, Appl
C 44	33.3	2.5	702	2	US-08-458-568A-3	Sequence 3, Appl
C 45	33.8	2.5	884	3	US-08-901-200A-11	Sequence 11, Appl

ALIGNMENTS

RESULT 1
US-08-533-669A-7/c
; Sequence 7, Application US/08533669A
; Patent No. 5834592
; GENERAL INFORMATION:
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: LEISHMANIA ANTIGENS FOR USE IN THE
; TITLE OF INVENTION: THERAPY AND DIAGNOSIS OF LEISHMANIASIS
; NUMBER OF SEQUENCES: 18
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: SEED AND BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/533,669A
; FILING DATE: 22-SEP-1995
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Makl, David J.
; REGISTRATION NUMBER: 30,392
; REFERENCE/DOCKET NUMBER: 210121.420
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1771 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..1698
; US-08-533-669A-7

Query Match 3.2%; Score 44; DB 3; Length 1771;
Best Local Similarity 45.8%; Pred. No. 0.0099;
Matches 152; Conservative 0; Mismatches 180; Indels 0; Gaps 0;

QY 376 cagctctctccagaggtggcagcaatggcccgctactaggggacctaagagcagcag 435


```
;
;
; NUMBER OF SEQUENCES: 19
; CURRENT APPLICATION DATA:
;   APPLICATION NUMBER: US/07/569,781
;   FILING DATE: 23-AUG-1990
; PRIOR APPLICATION DATA:
;   APPLICATION NUMBER: 464,499
;   FILING DATE: 12-JAN-1990
;   APPLICATION NUMBER: 405,605
;   FILING DATE: 11-SEP-1989
; SEQ ID NO:5:
;   LENGTH: 1879
; 5212296-5

Query Match          2.9%; Score 39.6; DB 7; Length 1879;
Best Local Similarity 56.0%; Pred. No. 0.16;
Matches 75; Conservative 0; Mismatches 59; Indels 0; Gaps 0;

QY 574 cagcagccctcgtctgccccccaccagcctgaagcctgacaccttcaggaccagctc 633
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 923 cagtacgcccctcgcgccgacccagcctcgtcccgcgcggtgaggaaactgctc 982

QY 634 caggagctctgcatcccccagacctggtcggggacttgccagcgtgtattgggagc 593
      | ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 983 cgtaactgcctatcgacatcgggcgccgctgcgcacggcggaacatcgaggtc 1042

QY 694 cagcggccctcct 707
      || ||| ||| ||| |||
Db 1043 gaggggcaactcat 1056

RESULT 5
US-08-242-677-1/c
; Sequence 1, Application US/08242677
; Patent No. 5671143
; GENERAL INFORMATION:
; APPLICANT: Gaynor, Richard B
; APPLICANT: Wu, Foon W.
; TITLE OF INVENTION: Cellular Nucleic Acid Binding Protein
; TITLE OF INVENTION: and Uses Thereof in regulating Gene Expression and in the
; TITLE OF INVENTION: Treatment of AIDS
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Arnold, White & Durkee
; STREET: P.O. Box 4433
; CITY: Houston
; STATE: TX
; COUNTRY: USA
; ZIP: 77210
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Mayfield, Denise L.
; REGISTRATION NUMBER: 33,732
; REFERENCE/DOCKET NUMBER: UTSD:401
; TELEPHONE: 713-787-1400
; TELEFAX: 713-789-2679
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 5173 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; FEATURE:
```

```
;
; NAME/KEY: CDS
; LOCATION: 1..4863
; US-08-242-677-1

Query Match          2.8%; Score 38.6; DB 1; Length 5173;
Best Local Similarity 48.4%; Pred. No. 0.48;
Matches 107; Conservative 0; Mismatches 114; Indels 0; Gaps 0;

QY 507 tggggtcagcgcccaacctgctgagcagcagctgggtgctccctgtgagcagcagcagcac 566
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 907 TGCACTGCGCCCGCCAGCTCCCGCCGACACTCCACCGCCCTCTGCAGCAGGTAGCGCGCTC 848

QY 567 actgtctcagcagcgccctctcttgcctcccccaccagcctgaagcctgacaccttcaggga 626
      || ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 847 GCTTCGGGTGTCAGGCGTCCGCTGCGCCAGCCGCCCTGCACCGTCTCCAGAGCGCC 788

QY 627 ceagctccagcagcctctgcatcccccagaaagactgctggtgggacttgccagcgtgtatt 686
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 787 AGCAGCGCGGCGGTCCGCGGCCGCGCTCGCGCGCGCGCGCGGTCTGCCCGCGGGCT 728

QY 687 tgggagccagcgggccccctcttgcattctgtgtgcccagcagc 727
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 727 CGGCAACAGCTTCTCGGCCAGGCGCTCAGGACCAGCAGC 687

RESULT 6
US-08-614-770A-1
; Sequence 1, Application US/08614770A
; Patent No. 5773267
; GENERAL INFORMATION:
; APPLICANT: WILLIAM R. JACOBS AND GRAHAM F. HATFULL
; TITLE OF INVENTION: D29 SHUTTLE PHASMIDS AND USES THEREOF
; NUMBER OF SEQUENCES: 1
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: AMSTER, ROTHSTEIN & EBENSTEIN
; STREET: 90 PARK AVENUE
; CITY: NEW YORK
; STATE: NEW YORK
; COUNTRY: U.S.A.
; ZIP: 10016
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5 INCH 1.44 Mb STORAGE
; MEDIUM TYPE: DISKETTE
; COMPUTER: IBM PC COMPATIBLE
; OPERATING SYSTEM: MS-DOS
; SOFTWARE: ASCII
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/614,770A
; FILING DATE: MARCH 7, 1996
; ATTORNEY/AGENT INFORMATION:
; NAME: ELIZABETH A. BOGOSIAN
; REGISTRATION NUMBER: 39,911
; REFERENCE/DOCKET NUMBER: 96700/402
; TELEPHONE: (212) 697-5995
; TELEFAX: (212) 286-0854 or 286-0082
; TELEX: TWX 710-581-4766
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 49272
; TYPE: NUCLEIC ACID
; STRANDEDNESS: DOUBLE
; TOPOLOGY: LINEAR
; MOLECULE TYPE: OLIGONUCLEOTIDE
; DESCRIPTION: NO
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: MYCOBACTERIOPHAGE
; INDIVIDUAL ISOLATE: D29
; US-08-614-770A-1
```

Query Match 2.8%; Score 38.4; DB 2; Length 49272;
Best Local Similarity 48.2%; Pred. No. 1.6;
Matches 108; Conservative 0; Mismatches 116; Indels 0; Gaps 0;

QY 523 ctgccgagagagagtggtggtccctgtggtgagcagatcacacactgtctccagcaggcc 582
DB 43074 CTGCGTGAACGCTGTGCTGCGCGGTGTGGTGGCGAGCTACGGGAAGTGGCTCAGCAGCAGC 43133

QY 583 ctccgtctgccccccaccagcctgaagcctgcacaccttcaggagaccagctccagagagctc 642
DB 43134 GTTGGTGTGACCCCTCCCTCGAGAGGGAGCGCGAACCCTCGCGCGGTGCACATGTAATC 43193

QY 643 tgcattcccccaagacctggtctggtgggacttgccagcgtgtatttggggagccagcggccc 702
DB 43194 GAGCAGCTTCCAGTACAGCGCGGACAGCGCGGAGGGTCCCGGTACATCGGTGAGGCGC 43253

QY 703 ctcttgattctgtgcccagcagcagggggcctggtgctgcgca 746
DB 43254 ATCATGTTGCCAGGATCAGCCGCTTGGCGCTGGCCGGTTCA 43297

RESULT 7
US-08-728-323A-1/c
; Sequence 1, Application US/08728323A
; Patent No. 5948676
; GENERAL INFORMATION:
; APPLICANT: Chang, Yuan
; APPLICANT: Bohenzky, Roy A.
; APPLICANT: Russo, James J.
; APPLICANT: Edelman, Isidore S.
; APPLICANT: Moore, Patrick S.
; TITLE OF INVENTION: Immediate Early Protein From Kaposi's
; TITLE OF INVENTION: Sarcoma-Associated Herpesvirus, DNA
; TITLE OF INVENTION: Encoding Same And Uses Thereof
; NUMBER OF SEQUENCES: 21
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Cooper & Dunham LLP
; STREET: 1185 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/728,323A
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: White, John P.
; REGISTRATION NUMBER: 28,678
; REFERENCE/DOCKET NUMBER: 0575/52268/JPW/MSC/SKS
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-391-0525
; TELEFAX: 212-391-0525
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3489 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..3489
US-08-728-323A-1

Query Match 2.8%; Score 37.8; DB 4; Length 3489;
Best Local Similarity 47.0%; Pred. No. 1.9;
Matches 117; Conservative 0; Mismatches 132; Indels 0; Gaps 0;

RESULT 8
US-08-770-379-20
; Sequence 20, Application US/08770379
; Patent No. 5849564
; GENERAL INFORMATION:
; APPLICANT: Chang, Yuan
; APPLICANT: Bohenzky, Roy A.
; APPLICANT: Russo, James J.
; APPLICANT: Edelman, Isidore S.
; APPLICANT: Moore, Patrick S.
; TITLE OF INVENTION: POLYPEPTIDES FROM KAPOSI'S SARCOMA-ASSOCIATED
; TITLE OF INVENTION: HERPESVIRUS, DNA ENCODING SAME AND USES THEREOF
; NUMBER OF SEQUENCES: 20
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Cooper & Dunham LLP
; STREET: 1185 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/770,379
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: White, John P.
; REGISTRATION NUMBER: 28,678
; REFERENCE/DOCKET NUMBER: 52342
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 278-0400
; TELEFAX: (212) 391-0525
; INFORMATION FOR SEQ ID NO: 20:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3207 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
US-08-770-379-20

Query Match 2.8%; Score 37.8; DB 3; Length 3207;
Best Local Similarity 47.0%; Pred. No. 1.9;
Matches 117; Conservative 0; Mismatches 132; Indels 0; Gaps 0;

Query Match	2.8%	Score 37.4:	DB 6;	Length 3435;
Best Local Similarity	46.4%;	Pred. No. 0.83;		
Matches 122;	Conservative 0;	Mismatches 141;	Indels 0;	Gaps 0;
Qy	430	agcattccagaaaagtctgtaagtttgggtcagcagcctgcagggggaggagactgccga	489	
Db	1390	AGCATGGTGGGCGCGTGAGATGGACATGCTGCAGGTGCTGCGGGGAGTACAAGCTC	1449	
Qy	490	gacgggtgtcagcgtctctggggtcagcgcacacctgcgcgagaggacagctgggtgccttg	549	
Db	1450	CGTCCCAACACGCTCAATCCGTGAGCTTCCATCTCTGGCGGACGAAAGAGGACGTG	1509	
Qy	550	ctggcaggatgcacacactgctccagcagcgcctccgtgtccccccaccacccctgaag	609	
Db	1510	CAGCAGACGATCATCACCGACCTGCAGAAATGGGAACACACAGACCCGCCCGCTGGCT	1569	
Qy	610	cttgacaccttcaggagaccagctccagagctctgcatcccccaagacctgggtcgggggac	669	
Db	1570	GTGTACTGCTGAAGATGCTTACCTGCCATCGCGGTGCTGTGAGCGGCTCATGTGTGCTG	1629	
Qy	670	ttggccagcgtggtattttggag	692	
Db	1630	GTGAAGCCGCTGGAGATGGCGAG	1652	

RESULT 11
 US-08-396-650-2
 Sequence 2, Application US/08396650
 Patent No. 5629204
 GENERAL INFORMATION:
 APPLICANT: HONJO, TASUKU
 APPLICANT: ISHIDA, YASUMASA
 APPLICANT: SHINOHARA, TAKASHI
 TITLE OF INVENTION: A NOVEL PEPTIDE RELATED TO HUMAN
 TITLE OF INVENTION: PROGRAMMED CELL DEATH AND DNA ENCODING IT
 NUMBER OF SEQUENCES: 4
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: SUGHRUE, MION, ZINN, MACPEAK & SEAS
 STREET: 2100 Pennsylvania Avenue, N.W.
 CITY: Washington
 STATE: D.C.
 COUNTRY: U.S.A.
 ZIP: 20037-3202
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC Compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patent in Release #1.0, Version #1.30
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/396,650
 FILING DATE: 01-MAR-1995
 CLASSIFICATION: 435
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: JP 55224/1994
 FILING DATE: 01-MAR-1994
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (202)293-7060
 TELEFAX: (202)293-7860
 TELEX: 6491103
 INFORMATION FOR SEQ ID NO: 2:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 864 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 MOLECULE TYPE: cDNA
 US-08-396-650-2

Query Match 2.6%; Score 35.8; DB 1; Length 864;
Best Local Similarity 49.7%; Pred. No. 1.2;

```

Matches 91; Conservative 0; Mismatches 92; Indels 0; Gaps 0
Qy 457 gtgtcagcagcctgcaggggaggagactgccgagacggtgtgcagcgtcttgggggtcagc 516
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 250 GAGGACCGCAGCCAGCGCGGCGAGCTGCGGCTTCGTGTGCACACAACGCGG 309
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 517 gccaaacctgcggaggagcagcactgggtgccctgtgtgcaggcatgcacacactctccag 576
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 310 CGTGACTTCCACATGAGCGTGGTCAGGGCCCGGGCGCAATGACAGCGGCACCTACTCTGT 369
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 577 caggccctccgtctgccccccaccagcctgaagcctgacaccttcaggggaccagctccag 636
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 370 GGGGCCATCTCCCTGGCCCCCAAGGCGCAGATCAAGAGAGAGAGAGAGAGAGAGAGAG 429
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

Qy 637 gag 639
    |||
Db 430 GTG 432

RESULT 12
US-08-768-626-2
; Sequence 2, Application US/08768626
; Patent No. 5698520
; GENERAL INFORMATION:
; APPLICANT: HONJO, TASUKU
; APPLICANT: ISHIDA, YASUMASA
; APPLICANT: SHINOHARA, TAKASHI
; TITLE OF INVENTION: A NOVEL PEPTIDE RELATED TO HUMAN
; TITLE OF INVENTION: PROGRAMMED CELL DEATH AND DNA ENCODING IT
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESS: SUGHRUE, MION, ZINN, MACPEAK & SEAS
; CITY: Washington
; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20037-3202
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08768,626
; FILING DATE: 01-MAR-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 55224/1994
; FILING DATE: 01-MAR-1994
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202)293-7060
; TELEFAX: (202)293-7860
; TELEX: 6491103
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 864 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
US-08-768-626-2

```

Query Match	2.6%	Score 35.8; DB 2; Length 864;
Best Local Similarity	49.7%	Pred. No. 1.2;

Qy 457 gtgtcagcagcctgcaggggaggactgccgagcgtgtgcagcgtcttggggtcagc 516
Db 250 GAGACCGCAGCCAGCCCGCCAGGACTGCCGTTCCGTTGCACAACTGCCACGGG 309
Qy 517 gccaacctgccggaggagcagcgggtgcctgctggcagcgatgcacacactgctccag 576

Db	310	CGTGACTTCACATGAGCGGTGTGTCAGGGCCCGGGCGCAATGACAGCGGCACCTACCTCTGT	369
Qy	577	caggccctcgtgtccccccaccagcctgaagcctgacaccttcagggaccagctccag	636
Db	370	GGGGCCATCTCCCTGGCCCCCAAGGGCGAGATCAAGAGAGCCTGCGGCGAGAGCTCAGG	429
Qy	637	gag	639
Db	430	GTG	432

```

RESULT 13
US-08-396-650-3
: Sequence 3, Application US/08396650
: Patent No. 5629204
: GENERAL INFORMATION:
: APPLICANT: HONJO, TASUKU
: APPLICANT: ISHIDA, YASUMASA
: APPLICANT: SHINOHARA, TAKASHI
: TITLE OF INVENTION: A NOVEL PEPTIDE RELATED TO HUMAN
: TITLE OF INVENTION: PROGRAMMED CELL DEATH AND DNA ENCODING IT
: NUMBER OF SEQUENCES: 4
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: SUGHRUE, MION, ZINN, MACPHEAK & SEAS
: STREET: 2100 Pennsylvania Avenue, N.W.
: CITY: Washington
: STATE: D.C.

```

```

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953
954
955
956
957
958
959
960
961
962
963
964
965
966
967
968
969
970
971
972
973
974
975
976
977
978
979
980
981
982
983
984
985
986
987
988
989
990
991
992
993
994
995
996
997
998
999
1000
1001
1002
1003
1004
1005
1006
1007
1008
1009
1010
1011
1012
1013
1014
1015
1016
1017
1018
1019
1020
1021
1022
1023
1024
1025
1026
1027
1028
1029
1030
1031
1032
1033
1034
1035
1036
1037
1038
1039
1040

```

```

1 / CONSTRUCTION: 433
2 /
3 / PRIOR APPLICATION DATA:
4 /
5 / APPLICATION NUMBER: JP 55224/1994
6 /
7 / FILING DATE: 01-MAR-1994
8 /
9 / TELECOMMUNICATION INFORMATION:
10 /
11 / TELEPHONE: (202)293-7060
12 /
13 / TELEFAX: (202)293-7860
14 /
15 / TELEX: 6491103
16 /
17 / INFORMATION FOR SEQ ID NO: 3:
18 /
19 / SEQUENCE CHARACTERISTICS:
20 /
21 / LENGTH: 921 base pairs
22 /
23 / TYPE: nucleic acid
24 /
25 / STRANDEDNESS: single
26 /
27 / TOPOLOGY: linear
28 /
29 / MOLECULE TYPE: CDNA
30 /
31 / US-08-3596-650-3

```

```
Query Match          2.68; Score 35.8; DB 1; Length 921;
Best Local Similarity 49.7%; pred. No. 1.2;
Matches 91; Conservative 0; Mismatches 92; Indels 0; Gaps 0;
```

457	ggtgtcagcagcctgcaggggagagactgcgagcgggtgcagcgtctctggggtcacg	516
QY		
274	gaggaccgcagccagccggccgaggaactgccgcttcctgtgcacacaaactgcccaacggg	333
Db		
517	gccaaacctgcgggagagcagctgggtgacctgtgcgagggcatgcacacactgctccag	576
QY		
334	cgtagacttccacatgatgacgtgtgcaggcccgccgacaaatgatcagcgcacactacctctct	393
Db		
577	cagcctctcgtctgccccccaccagcctgaagcctgcacaccttcaggagcagctccag	636
QY		
394	ggggccatcttcccttgcccccccaaggccgacgatcaaaagagagcctgcgggcagagctcagg	453
Db		
637	gag	639
QY		

Db 454 GTG 456

RESULT 14

```

US-08-396-650-4
; Sequence 4, Application US/08396650
; Patent No. 5629204
; GENERAL INFORMATION:
; APPLICANT: HONJO, TASUKU
; APPLICANT: ISHIDA, YASUMASA
; APPLICANT: SHINOHARA, TAKASHI
; TITLE OF INVENTION: A NOVEL PEPTIDE RELATED TO HUMAN
; TITLE OF INVENTION: PROGRAMMED CELL DEATH AND DNA ENCODING IT
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SUGHRUE, MIÖN, ZINN, MACPEAK & SEAS
; STREET: 2100 Pennsylvania Avenue, N.W.
; CITY: Washington
; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20037-3202
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/396,650
; FILING DATE: 01-MAR-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 55224/1994
; FILING DATE: 01-MAR-1994
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202)293-7060
; TELEFAX: (202)293-7860
; TELEX: 6491103
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 921 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; ORIGINAL SOURCE:
; ORGANISM: Homo sapiens
; CELL LINE: YTC3
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 25..888
; IDENTIFICATION METHOD: by similarity to some other pattern
; FEATURE:
; NAME/KEY: sig_peptide
; LOCATION: 25..84
; IDENTIFICATION METHOD: by similarity with known sequence
; IDENTIFICATION METHOD: or to an established consensus
; FEATURE:
; NAME/KEY: mat_peptide
; LOCATION: 85..888
; IDENTIFICATION METHOD: by similarity with known sequence
; IDENTIFICATION METHOD: or to an established consensus
US-08-396-650-4

```

Query Match 2.6%; Score 35.8; DB 1; Length 921;
Best Local Similarity 49.7%; Pred. NO. 1.2;

[illegible]

Db 334 CGTGACITCCACATGAGCGTGGTCCAGGCGCGCGCAATGACAGCGGCACCTACCTCTGT 393
Qy 577 caggccctccgtcgtcccccacagcctcgaagcctcacaccttcaggagaccagctccag 636
Db 394 GGGGCCATCTCCCTGGCCCCCAAGCGCGAGATCAAAGAGAGCGCTGCGGGCAGAGCTCAGG 453
Qy 637 gag 639
Db 454 GTG 456

Db 454 GTG 456

Search completed: May 15, 2000, 01:51:08
Job time: 10856 sec

RESULT 15
US-08-768-626-3
; Sequence 3, Application US/08768626
; Patent No. 5698520
; GENERAL INFORMATION:
; APPLICANT: HONJO, TASUKU
; APPLICANT: ISHIDA, YASUMASA
; APPLICANT: SHINOHARA, TAKASHI
; TITLE OF INVENTION: A NOVEL PEPTIDE RELATED TO HUMAN
; TITLE OF INVENTION: PROGRAMMED CELL DEATH AND DNA ENCODING IT
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SUGHRUE, MION, ZINN, MACPEAK & SEAS
; STREET: 2100 Pennsylvania Avenue, N.W.
; CITY: Washington
; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20037-3202
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/768,626
; FILING DATE: 01-MAR-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 55224/1994
; FILING DATE: 01-MAR-1994
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202)293-7060
; TELEFAX: (202)293-7860
; TELEX: 6491103
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 921 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
US-08-768-626-3

Query Match 2.6%; Score 35.8; DB 2; Length 921;
Best Local Similarity 49.7%; Pred. No. 1.2;
Matches 91; Conservative 0; Mismatches 92; Indels 0; Gaps 0;
Qy 457 gtggtcagcagcgtcaggggagagactccgagacggtgtgtcagcgtcttgggtcagc 516
Db 274 GAGGACCGAGCGAGCGCGCGGCGGACTCCCGTTCACACAACTGCCCAACGGG 333
Qy 517 gccacactccgagagcagcgtgggtccctgctggtcagcgatgcacacactgtccag 576
Db 334 CGTGACTTCCACATGAGCGGTGTCAGGGCCCGCGCAATGACAGGGGCACCTACCTCTGT 393
Qy 577 caggccctccgtcgtcccccacagcctgaagcctcacaccttcaggagaccagctccag 636
Db 394 GGGGCCATCTCCCTGGCCCCCAAGCGCGAGATCAAAGAGAGCGCTGCGGGCAGAGCTCAGG 453
Qy 637 gag 639

"

"

118 KLTDGSAHRFEVPIAKFQELRYSLVALVKEMAELEKKCKERLQD 224

Query Match 8.3%; Score 93.5; DB 1; Length 195;
Best Local Similarity 22.8%; Pred. No. 0.03;
Matches 33; Conservative 28; Mismatches 57; Indels 27; Gaps 5;

QY 85 LLAGMHTLQQALRLPTSLKPDFTF-RDQLQELCIPQDLVGLASVFGSRPLDLSVAQ 143
DB 71 LEAGHRADKSTL---STYEDCKFDRERIEFTCTEQNNKNSLEILGS----- 118

QY 144 QCGAWLPHVADFWRVDVAISTALARSQPSVLMQLKSDGSAYRFEVPTAKFQELRYS 203
DB 118 -IGRSLPHITDVSRLQYQIKTNQLHRMYRPAVLTLSVQN-----TDSPSYPEISFS 169

QY 204 VAL-----VLKEMADLEKRCERLQ 223
DB 170 CSMEQLQDLVGLKDKASKSLERATQ 194

RESULT 4
ID W06320
AC W06320
DT 07-FEB-1997 (first entry)
DE Human mitogen-activated protein kinase protein kinase 4-alpha.

118 KLTDGSAHRFEVPIAKFQELRYSLVALVKEMAELEKKCKERLQD 224

Query Match 8.3%; Score 93.5; DB 1; Length 195;
Best Local Similarity 22.8%; Pred. No. 0.03;
Matches 33; Conservative 28; Mismatches 57; Indels 27; Gaps 5;

QY 85 LLAGMHTLQQALRLPTSLKPDFTF-RDQLQELCIPQDLVGLASVFGSRPLDLSVAQ 143
DB 71 LEAGHRADKSTL---STYEDCKFDRERIEFTCTEQNNKNSLEILGS----- 118

QY 144 QCGAWLPHVADFWRVDVAISTALARSQPSVLMQLKSDGSAYRFEVPTAKFQELRYS 203
DB 118 -IGRSLPHITDVSRLQYQIKTNQLHRMYRPAVLTLSVQN-----TDSPSYPEISFS 169

QY 204 VAL-----VLKEMADLEKRCERLQ 223
DB 170 CSMEQLQDLVGLKDKASKSLERATQ 194

RESULT 3
ID Y07867
AC Y07867
DT 06-JUL-1999 (first entry)
DE Human secreted protein fragment encoded from gene 16.
KW Human; secreted protein; treatment; prevention; protein therapy; AIDS;
KW gene therapy; diagnosis; cancer; tumour; neurodegenerative disorder;

118 KLTDGSAHRFEVPIAKFQELRYSLVALVKEMAELEKKCKERLQD 224

Query Match 8.3%; Score 93.5; DB 1; Length 195;
Best Local Similarity 22.8%; Pred. No. 0.03;
Matches 33; Conservative 28; Mismatches 57; Indels 27; Gaps 5;

QY 85 LLAGMHTLQQALRLPTSLKPDFTF-RDQLQELCIPQDLVGLASVFGSRPLDLSVAQ 143
DB 71 LEAGHRADKSTL---STYEDCKFDRERIEFTCTEQNNKNSLEILGS----- 118

QY 144 QCGAWLPHVADFWRVDVAISTALARSQPSVLMQLKSDGSAYRFEVPTAKFQELRYS 203
DB 118 -IGRSLPHITDVSRLQYQIKTNQLHRMYRPAVLTLSVQN-----TDSPSYPEISFS 169

QY 204 VAL-----VLKEMADLEKRCERLQ 223
DB 170 CSMEQLQDLVGLKDKASKSLERATQ 194

RESULT 3
ID Y07867
AC Y07867
DT 06-JUL-1999 (first entry)
DE Human secreted protein fragment encoded from gene 16.
KW Human; secreted protein; treatment; prevention; protein therapy; AIDS;
KW gene therapy; diagnosis; cancer; tumour; neurodegenerative disorder;

118 KLTDGSAHRFEVPIAKFQELRYSLVALVKEMAELEKKCKERLQD 224

Query Match 8.3%; Score 93.5; DB 1; Length 195;
Best Local Similarity 22.8%; Pred. No. 0.03;
Matches 33; Conservative 28; Mismatches 57; Indels 27; Gaps 5;

QY 85 LLAGMHTLQQALRLPTSLKPDFTF-RDQLQELCIPQDLVGLASVFGSRPLDLSVAQ 143
DB 71 LEAGHRADKSTL---STYEDCKFDRERIEFTCTEQNNKNSLEILGS----- 118

QY 144 QCGAWLPHVADFWRVDVAISTALARSQPSVLMQLKSDGSAYRFEVPTAKFQELRYS 203
DB 118 -IGRSLPHITDVSRLQYQIKTNQLHRMYRPAVLTLSVQN-----TDSPSYPEISFS 169

QY 204 VAL-----VLKEMADLEKRCERLQ 223
DB 170 CSMEQLQDLVGLKDKASKSLERATQ 194

RESULT 3
ID Y07867
AC Y07867
DT 06-JUL-1999 (first entry)
DE Human secreted protein fragment encoded from gene 16.
KW Human; secreted protein; treatment; prevention; protein therapy; AIDS;
KW gene therapy; diagnosis; cancer; tumour; neurodegenerative disorder;

118 KLTDGSAHRFEVPIAKFQELRYSLVALVKEMAELEKKCKERLQD 224

Query Match 8.3%; Score 93.5; DB 1; Length 195;
Best Local Similarity 22.8%; Pred. No. 0.03;
Matches 33; Conservative 28; Mismatches 57; Indels 27; Gaps 5;

QY 85 LLAGMHTLQQALRLPTSLKPDFTF-RDQLQELCIPQDLVGLASVFGSRPLDLSVAQ 143
DB 71 LEAGHRADKSTL---STYEDCKFDRERIEFTCTEQNNKNSLEILGS----- 118

QY 144 QCGAWLPHVADFWRVDVAISTALARSQPSVLMQLKSDGSAYRFEVPTAKFQELRYS 203
DB 118 -IGRSLPHITDVSRLQYQIKTNQLHRMYRPAVLTLSVQN-----TDSPSYPEISFS 169

QY 204 VAL-----VLKEMADLEKRCERLQ 223
DB 170 CSMEQLQDLVGLKDKASKSLERATQ 194

RESULT 3
ID Y07867
AC Y07867
DT 06-JUL-1999 (first entry)
DE Human secreted protein fragment encoded from gene 16.
KW Human; secreted protein; treatment; prevention; protein therapy; AIDS;
KW gene therapy; diagnosis; cancer; tumour; neurodegenerative disorder;

KW Mitogen-activated protein kinase 4-alpha; MKK4-alpha; MAP;
 KW tyrosine kinase; signal transduction; cytokine; oncoprotein;
 OS stress; diagnosis; therapy.
 PN Homo sapiens.
 PD W09636642-A1.
 PD 21-NOV-1996.
 PF 26-JAN-1996; U01078.
 PR 19-MAY-1995; US-446083.
 PR 19-SEP-1995; US-530950.
 PA (DAVI//) DAVIS R J.
 PA (DERI//) DERIJARD B.
 PA (GUPT//) GUPTA S.
 PA (RAIN//) RAINGEAUD J.
 PI Davis RJ, Derijard B, Gupta S, Raingeaud J;
 DR WPI; 97-012035/01.
 DR N-PSDB; T43205.
 PT New mitogen activated protein kinase - useful for treating
 PT ischaemic heart disease, kidney failure etc., also for identifying
 PT modulators for treatment of similar conditions
 PS Claim 11; Fig 6; 104pp; English.
 CC Novel human mitogen activated protein kinase 4-alpha
 CC (MKK4-alpha) (W06320) has serine, threonine and tyrosine kinase
 CC activity, and mediates a signal transduction pathway that activates
 CC human mitogen-activated protein (MAP) kinases p38 and JNK. It is
 CC an alternatively spliced form of MKK4-gamma (W06322) and MKK4-beta
 CC (W06321) isoforms, which differ slightly at the NH2-terminus. MKK4
 CC is expressed in all tissues tested, but partic. in skeletal
 CC muscle. Recombinant MKK4 can be produced in transformed host cells.
 CC MKKs (W06318-22) can be used in the treatment of MKK-related
 CC disorders, e.g. ischaemic heart disease and kidney failure, to
 CC identify modulators of MKK activity, and to raise antibodies.
 SQ Sequence 363 AA;

Query Match 7.2%; Score 82; DB 1; Length 363;
 Best Local Similarity 20.2%; Pred. No. 1.1;
 Matches 59; Conservative 44; Mismatches 89; Indels 100; Gaps 13;

QY 3 AVGAATPYLHPGDSHSGRVSLGALPPEVAMARLLGLDRLSTFRKLKFFVSSLOGE 62
 DB 76 AYGSVNKMVHKP-----SG-----QIMAVKRIRSTVDEKEQKQLLMDLDDVVMRSS 120
 QY 63 DC-----RDG-----VORLGVSAN-----LPEQLGAL-LAGMHTL 92
 DB 121 DCPYIVQFYGALFREGDCWICMELMSTFDFKYKYVSVLDDVPEILGKITLATVKAL 180
 QY 93 --LQALRLPPTSLKPDTRFDQLQELCIPQDLVGLDLSVFGSQRPRLDSVAQOQAWL- 150
 DB 181 NHLKENLKIHRDIKPSN-----ILLDRSGNIKLCDFGIGQLVDSIAKTRDAGR 231
 QY 150 PHVADFR-----WRVDVAISTSAARSLOP---SVLMQL-----K 181
 DB 232 PYMAPERIDPSASRQGYDVRSDVWSLGITLYELATGRFPYKWNVSFDTQTVVKGDPQP 291
 QY 182 LSDGSAYRF-----EVPYAKFOELRYSVLVLKEMADLEKRC 218
 DB 292 LNSEEREFSFNFVNLCTKDESKRPYKELKHHPFILMYEERAVEVAC 343

RESULT 5

ID W97670
 AC W06321 standard; Protein; 363 AA.
 DT 10-MAY-1999 (first entry)
 DE Human mitogen activated protein kinase MKK4 alpha.
 KW MKK4 alpha; human; mitogen activated protein kinase kinase;
 KW MAP kinase kinase; signal transduction; inflammation; psoriasis;
 KW stress; diagnosis; therapy.
 OS Homo sapiens.
 PN W09902547-A1.
 PD 21-JAN-1999.
 PF 07-JUL-1998; U14101.
 PA (RAIN//) RAINGEAUD J.

PA (UTMA-) UNIV MASSACHUSETTS.
 PI Davis RJ, Tournier C, Whitmarsh A;
 DR WPI; 99-120771/10.
 DR N-PSDB; X07067.
 PT New isolated mitogen-activated protein kinase isoforms - used
 PT to develop products for treating e.g. inflammatory disorders,
 PT oxidative damage, proliferative disorders or autoimmune disorders
 PS Example 1; Page 132-133; 168pp; English.
 CC This polypeptide comprises human mitogen activated protein (MAP)
 CC kinase kinase 4 alpha (MKK4 alpha). Its sequence was deduced from
 CC a cDNA clone (see X07067) derived from human brain mRNA. MKK4
 CC includes 3 isoforms, termed MKK4 alpha, beta (see W97671) and gamma
 CC (see W97672), that vary slightly at the N-terminus. MKK4 is a p38
 CC MAP kinase kinase that also phosphorylates the JNK subgroup of MAP
 CC kinases. MKK3, MKK4, MKK6 and MKK7 (see W97664 and W97668-72),
 CC described in the invention, mediate the transduction of specific
 CC signals from the cell surface to the nucleus along specific
 CC pathways. They are useful for screening reagents which modulate
 CC MKK activity. Such agents can be used to prevent or treat
 CC MKK-mediated disorders, e.g. inflammation, oxidative damage or
 CC stress-related proliferative disorders, e.g. psoriasis, AIDS,
 CC malignancies of e.g. the skin, bone marrow, lung, liver, breast,
 CC gastrointestinal system and genito-urinary tract. Agents which
 CC inhibit the activity or expression of MKK inhibit cell growth or
 CC cause apoptosis. MKK7 polynucleotides (see X07059-64) and
 CC polypeptides (see W97662-67) are claimed.
 SQ Sequence 363 AA;

Query Match 7.2%; Score 82; DB 1; Length 363;
 Best Local Similarity 20.2%; Pred. No. 1.1;
 Matches 59; Conservative 44; Mismatches 89; Indels 100; Gaps 13;

QY 3 AVGAATPYLHPGDSHSGRVSLGALPPEVAMARLLGLDRLSTFRKLKFFVSSLOGE 62
 DB 76 AYGSVNKMVHKP-----SG-----QIMAVKRIRSTVDEKEQKQLLMDLDDVVMRSS 120
 QY 63 DC-----RDG-----VORLGVSAN-----LPEQLGAL-LAGMHTL 92
 DB 121 DCPYIVQFYGALFREGDCWICMELMSTFDFKYKYVSVLDDVPEILGKITLATVKAL 180
 QY 93 --LQALRLPPTSLKPDTRFDQLQELCIPQDLVGLDLSVFGSQRPRLDSVAQOQAWL- 150
 DB 181 NHLKENLKIHRDIKPSN-----ILLDRSGNIKLCDFGIGQLVDSIAKTRDAGR 231
 QY 150 PHVADFR-----WRVDVAISTSAARSLOP---SVLMQL-----K 181
 DB 232 PYMAPERIDPSASRQGYDVRSDVWSLGITLYELATGRFPYKWNVSFDTQTVVKGDPQP 291
 QY 182 LSDGSAYRF-----EVPYAKFOELRYSVLVLKEMADLEKRC 218
 DB 292 LNSEEREFSFNFVNLCTKDESKRPYKELKHHPFILMYEERAVEVAC 343

RESULT 6

ID W06321 standard; Protein; 393 AA.
 AC W06321;
 DT 07-FEB-1997 (first entry)
 DE Human mitogen-activated protein kinase kinase 4-beta.
 KW Mitogen-activated protein kinase kinase 4-beta; MKK4-beta; MAP;
 KW tyrosine kinase; signal transduction; cytokine; oncoprotein;
 KW stress; diagnosis; therapy.
 OS Homo sapiens.
 PN W09636642-A1.
 PD 21-NOV-1996.
 PF 26-JAN-1996; U01078.
 PR 19-MAY-1995; US-446083.
 PR 19-SEP-1995; US-530950.
 PA (DAVI//) DAVIS R J.
 PA (DERI//) DERIJARD B.
 PA (GUPT//) GUPTA S.
 PA (RAIN//) RAINGEAUD J.

RESULT	8	
W06322	W06322 standard; Protein; 399 AA.	
ID	AC W06322;	
DT	07-FEB-1997 (first entry)	
DE	Human mitogen-activated protein kinase 4-gamma.	
DE	Human mitogen-activated protein kinase 4-gamma; MKK4-gamma; MAP;	
KW	tyrosine kinase; signal transduction; cytokine; oncoprotein;	
KW	stress; diagnosis; therapy.	
KW	Homo sapiens.	
OS	Homo sapiens.	
PN	W06336642-A1.	
PD	21-NOV-1996.	
PF	26-JAN-1996; U01078.	
PR	19-MAY-1995; US-446083.	
PR	15-SEP-1995; US-530950.	
PA	(DAVI/) DAVIS R J	
PA	(DERI/) DERIJARD B.	
PA	(GUPY/) GUPTA S.	
PA	(RAIN/) RAINGEAUD J.	
PI	Davis RJ, Derijard B, Gupta S, Raingeaud J;	
DR	WPI: 97-012035/01.	
DR	N-PSDB: T43207.	
PT	New mitogen activated protein kinase kinase - useful for treating	
PT	ischemic heart disease, kidney failure etc., also for identifying	
PT	moleculatofs for treatment of similar conditions	
PT	Claim 19; Fig 8; 104pp; English.	
PS1	Novel human mitogen activated protein kinase kinase 4-gamma	
CC	(MKK4-gamma) (W06322) has serine, threonine and tyrosine kinase	
CC	activity, and mediates a signal transduction pathway that activates	
CC	human mitogen-activated protein (MAP) kinases p38 and JNK. Its	
CC	amino acid sequence was deduced from a cDNA clone (T43206) obtd.	
CC	from a human foetal brain library. Alternative splicing of the	

RESULT	7	
W97671		
AC	W97671 standard; Protein; 393 AA.	
DC	W97671.	
AT	10-MAY-1999 (first entry)	
DE	Human mitogen activated protein kinase kinase MKK4 beta.	
DE	MAPK beta; human; mitogen activated protein kinase kinase.	
KW	MAPK beta kinase; signal transduction; inflammation; psoriasis;	
KW	AIDS; cancer; apoptosis; therapy.	
QS	Homo sapiens.	
PN	W09902547-A1.	
PD	21-JAN-1999.	
PF	07-JUL-1998; U14101.	
PR	07-JUL-1997; US-888429.	
PA	(UYNA-) UNIV MASSACHUSETTS.	
PI	Davis RJ, Tournier C, Whitmarsh A;	
PI	WPI; 99-120771/10.	
DR	N-PSDB; X07068.	
PT	New isolated mitogen-activated protein kinase isoforms - used	
PT	to develop products for treating e.g. inflammatory disorders,	
PT	oxidative damage, proliferative disorders or autoimmune disorders	
PS	Example 1: Page 135-136; 168pp; English.	
CC	This polypeptide comprises human mitogen activated protein (MAP)	
CC	kinase kinase 4 beta (MKK4 beta). Its sequence was deduced from	
CC	a cDNA clone (see X07068) derived from human brain mRNA. MKK4	
CC	includes 3 isoforms, termed MKK4 alpha (see W97670), beta and gamma	
CC	(see W97672), that vary slightly at the N-terminus. MKK4 is a p38	

CC sequence yields MKK4 isoforms alpha (W06320) and beta (W06321).
 CC MKK4 is expressed in all tissues tested, but partic. in skeletal
 CC muscle. Recombinant MKK4 can be produced in transformed host cells.
 CC MKKs (W06318-22) can be used in the treatment of MKK-related
 CC disorders, e.g. ischaemic heart disease and kidney failure, to
 CC identify modulators of MKK activity, and to raise antibodies.
 CC Sequence 399 AA;

Query Match 7.2%; Score 82; DB 1; Length 399;
 Best Local Similarity 20.2%; Pred. No. 1.2;
 Matches 59; Conservative 44; Mismatches 89; Indels 100; Gaps 13;
 QY 3 AVGAATPYLHPGDSHSGRVSLGALQPPPEVAAMARLLGDLDRSTFRKLLKFFVSSLOGE 62
 DB 112 AYGSVNMVHKP-----SG-----QIMAVKRIRSTVDEKEQQLLMDLDVYMRSS 156
 QY 63 DC-----RDG-----VORLGVSAN-----LPEOLGAL-LAGMHTL 92
 DB 157 DCPYIVQFYGALFREGDCWICMELMSTSFDFKYKYVSVLDDVIPPEILKGLITLATVKAL 216
 QY 93 --LQALRLPPTSLKPDPTFRDLQELCIPQDLVGLASVFGSORPLDSDVAOQOAWL- 150
 DB 217 NHLKENLKIHRDIKPSN-----ILLDRSGNIKLCDFGIGSQLVDSIAKIRDACR 267
 QY 150 PHVADFR-----WRVDVAISTALSARLQ-----SVLMQL-----K 181
 DB 268 PYMAPERIDPSASROGYDVRSDWSLGITLYELATGRFPYPKWNVSFQDLTVVKGDPQ 327
 QY 182 LSDGSAYRF-----EVPTAKFOELRYSVALLKEMADLEKRC 218
 DB 328 LNSEEREFSFNFVNLCTNDESKRPYKELLKHPFILMYEERAVEVAC 379

RESULT 9

ID W97672 standard; Protein; 399 AA.
 AC W97672;
 DT 10-MAY-1999 (first entry)
 DE Human mitogen activated protein kinase MKK4 gamma.
 KW MKK4 gamma; human; mitogen activated protein kinase kinase;
 KW MAP kinase kinase; signal transduction; inflammation; psoriasis;
 KW AIDS; cancer; apoptosis; therapy.
 OS Homo sapiens.
 PN W09902547-A1.
 PD 21-JAN-1999.
 PF 07-JUL-1998; U14101.
 PR 07-JUL-1997; US-888429.
 PA (UYMA-) UNIV MASSACHUSETTS.
 PI Davis RJ, Tournier C, Whitmarsh A;
 DR WPI; 99-120771/10.
 DR N-PSDB; X07069.
 PT New isolated mitogen-activated protein kinase kinase isoforms - used
 PT to develop products for treating e.g. inflammatory disorders,
 PT oxidative damage, proliferative disorders or autoimmune disorders
 PS Example 1; Page 139; 168pp; English.
 CC This polypeptide comprises human mitogen activated protein (MAP)
 CC kinase kinase 4 gamma (MKK4 gamma). Its sequence was deduced from
 CC a cDNA clone (see X07067) derived from human brain mRNA. MKK4
 CC includes 3 isoforms, termed MKK4 alpha (see W97670), beta (see
 CC W97671) and gamma, that vary slightly at the N-terminus. MKK4 is a
 CC p38 MAP kinase kinase that also phosphorylates the JNK subgroup of
 CC MAP kinases. MKK3, MKK4, MKK6 and MKK7 (see W97664 and W97668-72),
 CC described in the invention, mediate the transduction of specific
 CC signals from the cell surface to the nucleus along specific
 CC pathways. They are useful for screening reagents which modulate
 CC MKK activity. Such agents can be used to prevent or treat
 CC MKK-mediated disorders, e.g. inflammation, oxidative damage or
 CC stress-related proliferative disorders, e.g. psoriasis, AIDS,
 CC malignancies of e.g. the skin, bone marrow, lung, liver, breast,
 CC gastrointestinal system and genito-urinary tract. Agents which
 CC inhibit the activity or expression of MKK inhibit cell growth or
 CC cause apoptosis. MKK7 polynucleotides (see X07059-64) and

CC polypeptides (see W97662-67) are claimed.
 SQ Sequence 399 AA;

Query Match 7.2%; Score 82; DB 1; Length 399;
 Best Local Similarity 20.2%; Pred. No. 1.2;
 Matches 59; Conservative 44; Mismatches 89; Indels 100; Gaps 13;
 QY 3 AVGAATPYLHPGDSHSGRVSLGALQPPPEVAAMARLLGDLDRSTFRKLLKFFVSSLOGE 62
 DB 112 AYGSVNMVHKP-----SG-----QIMAVKRIRSTVDEKEQQLLMDLDVYMRSS 156
 QY 63 DC-----RDG-----VORLGVSAN-----LPEOLGAL-LAGMHTL 92
 DB 157 DCPYIVQFYGALFREGDCWICMELMSTSFDFKYKYVSVLDDVIPPEILKGLITLATVKAL 216
 QY 93 --LQALRLPPTSLKPDPTFRDLQELCIPQDLVGLASVFGSORPLDSDVAOQOAWL- 150
 DB 217 NHLKENLKIHRDIKPSN-----ILLDRSGNIKLCDFGIGSQLVDSIAKIRDACR 267
 QY 150 PHVADFR-----WRVDVAISTALSARLQ-----SVLMQL-----K 181
 DB 268 PYMAPERIDPSASROGYDVRSDWSLGITLYELATGRFPYPKWNVSFQDLTVVKGDPQ 327
 QY 182 LSDGSAYRF-----EVPTAKFOELRYSVALLKEMADLEKRC 218
 DB 328 LNSEEREFSFNFVNLCTNDESKRPYKELLKHPFILMYEERAVEVAC 379

RESULT 10

ID R15345 standard; Protein; 997 AA.
 AC R15345;
 DT 25-MAR-1992 (first entry)
 DE Mouse bullous pemphigoid antigen protein.
 KW dermatological syndrome; blisters; murine.
 OS Mus musculus.
 PN J03206885-A.
 PD 10-SEP-1991.
 PF 09-APR-1990; 093585.
 PR 31-OCT-1989; JP-283947.
 PR 09-APR-1990; JP-093585.
 PA (TOFU) TONEN CORP.
 DR WPI; 91-356566/49.
 DR N-PSDB; Q15019.
 PT DNA which encodes mouse derived bullous pemphigoid antigen protein -
 PT contg. base sequence which encodes polypeptide composed of specified
 PT 996 aminoacid sequence
 PS Disclosure; Fig 1; 23pp; Japanese.
 CC The polypeptide is recognised by auto-antibodies in the serum of a
 CC bullous pemphigoid patient and by a human anti-basement membrane zone
 CC monoclonal antibody. It can be used for diagnosis of bullous
 CC pemphigoid. See also Q15020.
 SQ Sequence 997 AA;

Query Match 7.2%; Score 82; DB 1; Length 997;
 Best Local Similarity 22.4%; Pred. No. 4.2;
 Matches 59; Conservative 30; Mismatches 106; Indels 68; Gaps 11;
 QY 2 SAVGAATPYLHPGDSHSGRVSL-----GAQLPPEVAA 35
 DB 228 SAFTAGRECHHPAEISPGNSGHLNKLRLPLSRWTOEPHTEKWPRAAEQLPKEV-- 286
 QY 36 MARLLG-DLDRSTFRKLLKFFVSSLOGE-----DCRDGVORLGVSANLPEQALLAGM 89
 DB 286 QFRQPGAPLDRESSQPCYSYFSOTSTELQITFDDKNPIITR-----LSEL 330
 QY 90 HTLLQQLR--LPPTSLKPDPTFRDLQELCIPQDLV-----GDLASVFGSORPLDSDVA 142
 DB 331 ETMREQALHPSRPVTVYQDDKLERELVKLLTPLEIAKNKCGMHTVTTLKQEKRLGSSA 390
 QY 143 QQQGAWLPHVADFRWRVDVAISTALSARLQPSVLMLQLSDGSAYRFEVPTAKFOELRY 202

Mon May 15 14:56:40 2000

Db 391 ---GGWLEGC-----RTSGGLKGDFLKSKVEPEASPSLDLQACSVRDE---EFQOGLRH 441

Qy 203 SVA---LVLKEMADLEKRCERRL 222

Db 442 TVTGRQLVEAKLLDNRIVEQLRL 464

RESULT 11

R51884

ID R51884 standard; Protein; 384 AA.

AC R51884; 1994 (first entry)

DE KAPA synthase encoded by bioF gene in plasmid pB030A-15/9.

KW Biotin; expression; enterobacteria; vitamin H; synthesis;

KW plasmid; pB030A-15/9; bioB; bioC; bioD; bioA;

KW promoter; ptac; biotin synthase; KAPA synthase;

KW 8-amino-7-oxononanoate synthase; pimeloyl-CoA; DTB synthase;

KW dehydrobiotin synthase; DAPA synthase;

KW S-adenosyl-L-methionine; 8-amino-7-oxononanoate aminotransferase;

KW seborrhea; dermatitis.

OS Escherichia coli DSM498.

PN WO9408023-A.

PD 14-APR-1994.

PF 01-OCT-1993; E02688.

PR 02-OCT-1992; CH-003124.

PR 15-JUL-1993; CH-002134.

PA (LONZ) LONZA AG.

PI Birch O, Brass J, Fuhrmann M, Shaw N;

DR WPI; 94-135587/16.

DR N-PSDB; 062386.

PT Biotechnological biotin prodn. using enterobacterial biotin-gene

PT providing vitamin H in high yield

PS Disclosure: Fig 6, Page 47-55 and 60-65; 92pp; German.

CC Plasmid pB030A-15/9 contains the bioB, bioC, bioD and bioA

CC genes responsible for biosynthesis of biotin, arranged in a

CC transcription unit. Microorganisms contg. these DNA fragments or

CC plasmids may be used in the prodn. of biotin. Biotin (Vitamin H)

CC may prevent seborrhea, dermatitis, loss of appetite and tiredness.

CC Sequence 384 AA;

Qy 6 AATPYL---HHP-----GDSHSGRVSFLGAQLPPEVAAARLLGLDRST 47

Db 45 SSNDYLGSLHHFQIIRAWOQGAEQFGIGSGGSHVSVVHQALEEAEWLG----- 99

Qy 48 FRLLKF-----VVSLOGEDCRGVRLGVSANLPEQLGALLAGMHTLLQALR 98

Db 99 YSRALLFISGFANQAVIAAMKEDRIADRLS-----HASLLEAAS 141

Qy 99 LPPTSLLKPTDFD-----OLQELCIPQDLVGLASVVF---GSORPL--LDSVAQOQGA 147

Db 142 LSPSLRRFAHNDVTHLARLLASPCPGQOMV--VTEGVFSMDGDSAPLAEIOOVTOOHNG 199

Qy 148 WLPVHADFRRVVDVAISTALA-----RSLOPSVLM----- 179

Db 200 WL-----WVDDAGTGVIGEGRGSCWLVKVPKELLVTFVGKGFVSGAAVLCSSTV 251

Qy 179 ---QLKLSGCSAYRFEVPTAKFOELRYSVLVKEMADLEK 216

Db 252 ADYLLQFARHLIYSTMPPAQAQALRASLAVIRSDGEGDARR 292

RESULT 13

R99975

ID R99975 standard; Protein; 335 AA.

AC R99975;

DT 29-APR-1997 (first entry)

DE Pseudomonas sp. SD705 (FERM BP-4772) lipase.

KW Lipase; detergent; food processing; paper; oil; manufacture

KW vector; recombinant production; transformed host; SD705;

KW Pseudomonas; alcaligenes; SD702; Bacillus; NKS-21.

OS Pseudomonas sp.

PN WO9627002-A1.

PD 06-SEP-1996.

PF 23-FEB-1996; J00426.

PR 27-FEB-1995; JP-038527.

PA (SHOW) SHOWA DENKO KK.

PI Ohno K, Sasuga J, Takada H, Yoneda T;

DR WPI; 96-412770/41.

DR N-PSDB; T36695.

PT Gene encoding lipase isolated from Pseudomonas SD705 - useful

PT industrially in detergents, food processing and paper and oil

PT manufacture

PS Claim 3; Pages 32-34; 52pp; Japanese.

Qy 6 AATPYL---HHP-----GDSHSGRVSFLGAQLPPEVAAARLLGLDRST 47

Db 45 SSNDYLGSLHHFQIIRAWOQGAEQFGIGSGGSHVSVVHQALEEAEWLG----- 99

Qy 48 FRLLKF-----VVSLOGEDCRGVRLGVSANLPEQLGALLAGMHTLLQALR 98

Db 99 YSRALLFISGFANQAVIAAMKEDRIADRLS-----HASLLEAAS 141

Qy 99 LPPTSLLKPTDFD-----OLQELCIPQDLVGLASVVF---GSORPL--LDSVAQOQGA 147

Db 142 LSPSLRRFAHNDVTHLARLLASPCPGQOMV--VTEGVFSMDGDSAPLAEIOOVTOOHNG 199

Qy 148 WLPVHADFRRVVDVAISTALA-----RSLOPSVLM----- 179

Db 200 WL-----WVDDAGTGVIGEGRGSCWLVKVPKELLVTFVGKGFVSGAAVLCSSTV 251

Qy 179 ---QLKLSGCSAYRFEVPTAKFOELRYSVLVKEMADLEK 216

Db 252 ADYLLQFARHLIYSTMPPAQAQALRASLAVIRSDGEGDARR 292

RESULT 12

W73904

ID W73904 standard; Protein; 384 AA.

AC W73904;

DT 12-APR-1999 (first entry)

DE -E. coli KAP synthetase.

" "

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run On: May 15, 2000, 03:22:30 ; Search time 35.18 Seconds
(without alignments)
92.094 Million cell updates/sec

Title: US-09-223-796-4
Sequence: 1 MSAGVGAATPYLHPGDSHSG.....ALVLKEMADLEKRCERRLQD 224
Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 143561 seqs, 14463640 residues

Total number of hits satisfying chosen parameters: 143561

Minimum DB seq length: 0

Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database :

- 1: /cgn2_6/ptodata/2/iaa/5A_COMB.pep.*
- 2: /cgn2_6/ptodata/2/iaa/5B_COMB.pep.*
- 3: /cgn2_6/ptodata/2/iaa/6_COMB.pep.*
- 4: /cgn2_6/ptodata/2/iaa/PCTUS_COMB.pep.*
- 5: /cgn2_6/ptodata/2/iaa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	96	8.5	195	2	US-08-822-260-3
2	93.5	8.3	195	2	US-08-822-260-1
3	82	7.2	363	1	US-08-530-950-6
4	82	7.2	399	1	US-08-530-950-10
5	82	7.2	399	2	US-08-874-186-92
6	81.5	7.2	384	2	US-08-401-068-2
7	81.5	7.2	384	2	US-08-846-338-2
8	81	7.2	335	2	US-08-875-062-3
9	79	7.0	393	1	US-08-530-950-8
10	77	6.8	340	1	US-08-606-888A-5
11	71.5	6.3	1103	1	US-08-455-543A-53
12	71.5	6.3	1103	2	US-08-223-305C-53
13	70.5	6.2	489	3	US-08-369-822C-27
14	70.5	6.2	3218	1	US-08-764-100-27
15	70	6.2	1106	1	US-08-435-675B-5
16	70	6.2	1106	1	US-08-336-257A-8
17	69.5	6.1	306	1	US-08-328-322-2
18	69.5	6.1	915	1	US-08-328-322-5
19	69	6.1	395	1	US-08-318-947A-18
20	69	6.1	395	2	US-08-795-303-18
21	69	6.1	446	3	US-08-956-254-2
22	69	6.1	448	2	US-09-015-815-1
23	69	6.1	1086	5	5386025-8
24	69	6.1	2476	1	US-08-276-967-2
25	69	6.1	3075	2	US-08-460-309-5
26	69	6.1	3075	2	US-08-125-077-5
27	67.5	6.0	551	2	US-08-436-771-9
28	67.5	6.0	551	2	US-08-434-998-9
29	67.5	6.0	551	2	US-08-487-797-9

30	67.5	6.0	551	4	PCT-US95-02058-9
31	67	5.9	1100	1	US-08-357-598-11
32	67	5.9	1100	2	US-09-003-289-11
33	67	5.9	1100	4	PCT-US95-16435-11
34	66.5	5.9	393	2	US-08-559-303B-74
35	66.5	5.9	560	2	US-08-095-728B-6
36	66.5	5.9	560	4	PCT-US92-02320A-6
37	66.5	5.9	1417	2	US-08-559-303B-78
38	66.5	5.9	2629	2	US-08-751-189-4
39	66.5	5.9	2629	2	US-09-060-836-4
40	66	5.8	15281	2	US-08-471-119A-2
41	65.5	5.8	387	5	5240849-5
42	65.5	5.8	401	2	US-08-591-079-4
43	65.5	5.8	494	1	US-08-447-500-2
44	65.5	5.8	494	1	US-08-454-097-2
45	65.5	5.8	494	1	US-08-447-408-2

ALIGNMENTS

RESULT 1
US-08-822-260-3
; Sequence 3, Application US/08822260
; Patent No. 5830660
; GENERAL INFORMATION:
; APPLICANT: Hillman, Jennifer L.
; APPLICANT: Goli, Surya K.
; TITLE OF INVENTION: NOVEL TUMORIGENESIS PROTEIN
; NUMBER OF SEQUENCES: 3
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: Fastseq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/822,260
; FILING DATE: Herewith
; CLASSIFICATION: 514
; PRIOR APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Billings, Lucy J.
; REGISTRATION NUMBER: 36,749
; REFERENCE/DOCKET NUMBER: PF-0247 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-855-0555
; TELEFAX: 415-845-4166
; TELEX:
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 195 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: GenBank
; CLONE: 265569
US-08-822-260-3

Query Match 8.5%; Score 96; DB 2; Length 195;
Best Local Similarity 20.8%; Pred. No. 0.0017;
Matches 43; Conservative 32; Mismatches 78; Indels 54; Gaps 8;

QY 41 GDLRSTERKLLKFVSVSLQDCRQVGRVGSANLPPEOLGALLAGMHTLQOALRLP 100
Db 18 GSFONAFALLRAAFQSL-----LDARADALD-----HPYLKQ---ID 55
QY 101 PTLK-----PDTFRDQLQELCIPQDLVGLASVVFSGRPLDVS 141
Db 56 PVVLKCHAAATCILEAGKHQVDKSTLSTYLEDCKFDRERI-ELFCTEYQNNKNSLETL 114
QY 142 AQOQGANLPHVADFRVVDVAISTALSARLQPSVLMQKL--SDGSAY---RFEVPTAK 196
Db 115 LGSIGRSLPHITDVSRLQYQIKTNQLHMYRPAVLTSLVQN-----TDSYSPISFS 169
QY 197 FOELRYVALVLEKADLEKRCERLQ 223
Db 175 LODL-----VGKLDASKSLERATQ 194

RESULT 2
US-08-822-260-1
; Sequence 1, Application US/08822260
; Patent No. 5830660
; GENERAL INFORMATION:
; APPLICANT: Hillman, Jennifer L.
; APPLICANT: Goli, Surya K.
; TITLE OF INVENTION: NOVEL TUMORIGENESIS PROTEIN
; NUMBER OF SEQUENCES: 3
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/822.260
; FILING DATE: Herewith
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Billings, Lucy J.
; REGISTRATION NUMBER: 36,749
; REFERENCE/DOCKET NUMBER: PF-0247 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-855-0555
; TELEFAX: 415-845-4166
; TELEX:
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 195 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: UTRNOT02
; CLONE: 2267574
US-08-822-260-1

Query Match 8.3%; Score 93.5; DB 2; Length 195;
Best Local Similarity 22.8%; Pred. No. 0.0034;
Matches 33; Conservative 28; Mismatches 57; Indels 27; Gaps 5;
QY 85 LLAGMHTLQOALRLPPTSLKPTDF-RDQLQELCIPQDLVGLASVVFSGRPLDVSVAQ 143
Db 71 LEAGKHRADSTL---STYLEDCKFDRERIELEFCTEYQNNKNSLEILGS----- 118

QY 144 QOQGANLPHVADFRVVDVAISTALSARLQPSVLMQKLSDGSAYRFEVPTAKFQELRYS 203
Db 118 -IGRSLPHITDVSRLQYQIKTNQLHMYRPAVLTSLVQN-----TDSYSPISFS 169
QY 204 VAL-----VLKADLEKRCERLQ 223
Db 170 CSMEQLQDLVGLKLDASKSLERATQ 194

RESULT 3
US-08-530-950-6
; Sequence 6, Application US/08530950
; Patent No. 5736381
; GENERAL INFORMATION:
; APPLICANT: Davis, Roger J.
; APPLICANT: Raugeaud, Joel
; APPLICANT: Gupta, Shashi
; APPLICANT: Derijard, Benoit
; TITLE OF INVENTION: CYTOKINE-, STRESS-, AND
; TITLE OF INVENTION: ONCOPROTEIN-ACTIVATED HUMAN PROTEIN KINASE
; TITLE OF INVENTION: KINASES
; NUMBER OF SEQUENCES: 16
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson P.C.
; STREET: 225 Franklin Street
; CITY: Boston
; STATE: MA
; COUNTRY: USA
; ZIP: 02110-2804
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/530,950
; FILING DATE: 19-SEP-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Fasse, J. Peter
; REGISTRATION NUMBER: 32,983
; REFERENCE/DOCKET NUMBER: 07917/010001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617/542-5070
; TELEFAX: 617/542-8906
; TELEX: 200154
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 363 amino acids
; TYPE: amino acid
; STRANDEDNESS: not relevant
; TOPOLOGY: linear
US-08-530-950-6

Query Match 7.2%; Score 82; DB 1; Length 363;
Best Local Similarity 20.2%; Pred. No. 0.2;
Matches 59; Conservative 44; Mismatches 89; Indels 100; Gaps 13;
QY 3 AVGAATPYLHHGDSHSGRVSFLGAQLPPEVAAMARLLGLDLDSTFRKLLKFVYVSSLQGE 62
Db 76 AVGSYKMKVHP-----SG-----QIMAYKRSTVDEKEQQLLMDLVVRRSS 120
QY 63 DC-----RDG-----VQRLGVSAN-----LPESOLGAL-LAGMHTL 92
Db 121 DCPYIVQVYGALFREGDCWICMELMSTSFDFKYKYVSVLDDVPIPEELGKITLATVKAL 180
QY 93 --LQOALRLPSTLKPDTFRQLQELCIPQDLVGLASVVFSGRPLDVSVAQOQAWL- 150
Db 181 NLKENLAKIHRDIKPSN-----LLDRSGNIKLCDFGSGQLVDSIAKTRDAGCR 231
QY 150 PHVADFR-----WRVDVAISTALSARLQP---SVLMOL-----K 181

Db 232 PYMAPERIDPSASRQGYDRSDVMSGLITLYELATGREPYKWNVSFDTQTQVKGDPPQ 291

Qy 182 LSDGSAYRF-----EVPYAKFQLRYSVLVLMKEMADLEKRC 218
||| | | | | : : : | : | :

Db 292 LNSNEERFSPSFNFVNCLTKDESKRPKYKELLKHPIFLMTEERAEEVAC 343

```

RESULT      4
US-08-530-950-10
; Sequence 10, Application US/08530950
; Patent No. 5736381
; GENERAL INFORMATION:
; APPLICANT: Davis, Roger J.
; APPLICANT: Raingeaud, Joel
; APPLICANT: Gupta, Shashi
; APPLICANT: Derjard, Benoit
; TITLE OF INVENTION: CYTOKINE-, STRESS-, AND
; TITLE OF INVENTION: ONCOPROTEIN-ACTIVATED HUMAN PROTEIN KINASE
; TITLE OF INVENTION: KINASES
; NUMBER OF SEQUENCES: 16
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson P.C.

```

COUNTRY: USA
 ZIP: 02110-2804
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: PatentIn Release #1.0, Version #1.30
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/530,950
 FILING DATE: 19-SEP-1995
 CLASSIFICATION: 435
 ATTORNEY/AGENT INFORMATION:
 NAME: Fasse, J. Peter
 REGISTRATION NUMBER: 32,983
 REFERENCE/DOCKET NUMBER: 07917/010001
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 617/542-5070
 TELEFAX: 617/542-8906
 TELEX: 200154
 INFORMATION FOR SEQ ID NO: 10:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 399 amino acids
 TYPE: amino acid
 STRANDEDNESS: not relevant
 TOPOLOGY: linear
 US-08-530-950-10

Query Match 7.28; Score 82; DB 1; Length 399;
Best Local Similarity 20.2%; Pred. No. 0.23;
Matches 59; Conservative 44; Mismatches 89; Indels 100; Gaps 13;

Qy		3	AVGAATPYLLHFGDSHGSRVSLFQAQLPPEVAAMARLLGLDLDRSTFRKLLKTFWVSLSLOE	62
			: :	
Db		112	AGSVNKKYHKP-----SG-----QIMAVKRIRSTVDEKEQQLMLDLDVVYMRSS	156
			: :	
Qy		63	DC-----RQG-----VORLGSAN-----LPEQLGAL-LAGMHTL	92
			: :	
Db		157	DCPYIVQFYGFALFREGDCHWICHELMTSFDKYKVYSVLDDVIPFEELGKITLATVRAL	216
			: :	
Qy		93	--LQALRLPTSLKPDTERDOLQEILCPQDLVGDLASVVFQSQRPLDSVAQQQAWL-	150
			: :	
Db		217	NHLKENLKIIHRDIKPSN-----ILLDSRGNIKLCDFGISGLVDSIATKTRDAGR	267
			: :	
Qy		150	PHVAOFR-----WRVDAISTARSLSQP---SVLMQL-----K	181
			: :	
Db		268	PYMAERIDPSASROGYDVRSVDWSLGITLYELATGRPPYPKNWSVFDTQTQVVKGDPPQ	327

QY	182	LSDGSAYRF-----EVPTAKFQELRYSVALVLKEMADLEKRC	218
		:	
Dd	328	LSNSEEREFSPSFINFVNLOITKDESKRPKYKELLKHHPFTILMYEERAVEVAC	379

RESULT 5
US-08-874-186-92
; Sequence 92, Application US/08874186
; Patent No. 5989885
; GENERAL INFORMATION:
; APPLICANT: Teng, David H-F.
; APPLICANT: Tavligian, Sean V.
; APPLICANT: Perry III, William L.
; APPLICANT: Skolnick, Mark H.
; TITLE OF INVENTION: SPECIFIC MUTATIONS OF MAP KINASE KINASE
; TITLE OF INVENTION: 4 (MKK4) IN HUMAN TUMOR CELL LINES IDENTIFY IT AS A TUMOR
; TITLE OF INVENTION: SUPPRESSOR IN VARIOUS TYPES OF CANCER
; NUMBER OF SEQUENCES: 96
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Venable, Baetjer, Howard & Civiletti, LLP
; STREET: 1201 New York Avenue, N.W., Suite 1000
; CITY: Washington
; STATE: DC
; COUNTRY: U.S.A.

```

1 215 20003
2
3  COMPUTER READABLE FORM:
4
5  MEDIUM TYPE: Floppy disk
6
7  COMPUTER: IBM PC Compatible
8
9  OPERATING SYSTEM: PC-DOS/MS-DOS
10
11 SOFTWARE: PatentIn Release #1.0, Version #1.30
12
13 CURRENT APPLICATION DATA:
14
15 APPLICATION NUMBER: US/08/874,186
16
17 FILING DATE:
18
19 CLASSIFICATION: 435
20
21 PRIOR APPLICATION DATA:
22
23 APPLICATION NUMBER: US 08/782,482
24
25 FILING DATE: 10-JAN-1997
26
27 ATTORNEY/AGENT INFORMATION:
28
29 NAME: Saxe, Stephen A.
30
31 REGISTRATION NUMBER: 38,609
32
33 REFERENCE/DOCKET NUMBER: 24884-121392-01
34
35 TELECOMMUNICATION INFORMATION:
36
37 TELEPHONE: 202-962-4948
38
39 TELEFAX: 202-962-8300
40
41 INFORMATION FOR SEQ ID NO: 92:
42
43 SEQUENCE CHARACTERISTICS:
44
45 LENGTH: 399 amino acids
46
47 TYPE: amino acid
48
49 TOPOLOGY: linear
50
51 MOLECULE TYPE: protein
52
53 US-08-874-186-92

```

Query Match 7.2%; Score 82; DB 2; Length 399;
Best Local Similarity 20.2%; Pred. No. 0.23;
Matches 59; Conservative 44; Mismatches 89; Indels 100; Gaps 13;

QY	3	AVGAATPYLHHPGDSHSGRVSYFGALPPEVAAARLLGDLDRSTFRKLLKFVYSSLOQE	62
Ddb	112	AYGSVNKHVP-----SG-----OIMAVRKIRSTVDEKEQQLMLDLDVVMRSS	156
QY	63	DC-----RDG---VORLGVSAN-----LPBOLGAL-LAGMHTL	92
Ddb	157	DCPIYIVQFYGALFREGDCWICMEI MSTDFKYKYVYGVLDVDVIPETILKITITVKAL	216
QY	93,	--LQAALRPPTSLKPDTFRDQLQELCIPQDLVGVLASVWFSGSRPLDSDVAQQQAWL-	150
Ddb	217	NHLKENLKITHRIKPSN-----ILLDRSGNIKCDFGISGLQVDSIAKTDRACGR	267
QY	150	PHVADFR-----WRDYDAISTSAIARSIQP---SVLMQL-----K	181
Ddb	268	PYMAPERIDPSAQGYDVNSDWSIGITLYIELATGREPYPKWNSVFEDLTQVWKGDPPQ	327

```

QY 182 LSDGSAYRF-----EVPTAKFOELYSVALVLKEMADLEKFC 210
DB 328 LSNSEEREFSPFINFVNCLTKDESKRPKYKELLKHPFILMYEERAVEVAC 379

RESULT 6
US-08-401-068-2
; Sequence 2, Application US/08401068
; Patent No. 5859335
; GENERAL INFORMATION:
; APPLICANT: Patton, David
; TITLE OF INVENTION: Enhanced Biotin Biosynthesis in Plant Tissue
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: CIBA-GEIGY Corporation
; STREET: 7 Skyline Drive
; CITY: Hawthorne
; STATE: NY
; COUNTRY: USA
; ZIP: 10532
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30B
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/401,068
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/351,970
; FILING DATE: 08-DEC-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Elmer, James Scott
; REGISTRATION NUMBER: 36,129
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 919-541-8614
; TELEFAX: 919-541-8689
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 384 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-401-068-2

```

```

Query Match      7.23;  Score 81.5;  DB 2;  Length 384;
Best Local Similarity 19.95;  Pred. No. 0.25;
Matches 56;  Conservative 32;  Mismatches 90;  Indels 103;  Gaps 12;

Qy      6  AATPYL---HHP-----GDSHSGRVSYFLGAQLPPEVAAMARLLGLDRST 47
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      45  SSNDYLGLSLHHPQITIRANQOQAEQFGICSGSGHVSQYVVHQALEEELAEWLG----- 99
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy      48  FRKLKF-----VVSSLQGEDCRDGVORLGVYSANLPEQLGALLAGMHTLQQALR 98
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      99  YSRALLFISGFAANOAVTAAAMMAKEDRIAADRLS-----HASTLEAAS 141
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy      99  LPPTSLKPDTPRD-----QLQELCIPDLVGDLASVVF--GSRPL--LDSVAQOQGA 147
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      142  LSPQLRFRFANDVTHLARLLASPCGQOMV--VTEGVFSGMDGSAPLAETIQVTVQOHNG 199
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy      148  WLPHVDFRFRVDVAISTASALA-----RSLOPSVLM----- 179
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      200  WL-----MYDDAHGTGVIGEQGRGCSWLQKVAPPELLVTFGFGVSGAAVLCSTV 251
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy      179  ---QLKLSGDSAYREVEVTAKFQELRYSVALKEMADLEK 216
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      252  ADYLQFAHLLIYSTMPPPAQALRALSLAVIRSDGDRR 292

```

```

US-08-846-338-2
; Sequence 2, Application US/08/46338
; Patent No. 5869719
; GENERAL INFORMATION:
; APPLICANT: Patton, David
; TITLE OF INVENTION: Transgenic Plants Having Increased Biotin Content
; NUMBER OF SEQUENCES: 20
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: No. 5869719artis Corporation
; STREET: 520 White Plains Road, P.O. Box 2005
; CITY: Tarrytown
; STATE: NY
; COUNTRY: USA
; ZIP: 10591
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30B
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/846,338
; FILING DATE:
; CLASSIFICATION: 800
; ATTORNEY/AGENT INFORMATION:
; NAME: Meigs, J. Timothy
; REGISTRATION NUMBER: 38,241
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 919-541-8587
; TELEFAX: 919-541-8689
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 384 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-846-338-2

Query Match 7.2%; Score 81.5; DB 2; Length 384;
Best Local Similarity 19.9%; Pred. No. 0.25;
Matches 56; Conservative 32; Mismatches 90; Indels 103; Gaps

Qy 6 AATPYL---HHP-----GDSGRYSFGLAQLPPEVAMARLLGDLDRST 47
   : : || || | | | | | | | | | | | | | | | | | | | |
Db 45 SSNDILGLSHHPOIRAWOOGAEQFGIGSGSGSYVWHALEELAEWLG----- 99
   : : || || | | | | | | | | | | | | | | | | | | | |

Qy 48 FRKLKK-----VVSQGEDCRDGVRLGYSANLPEELGALLAGMHTLIQQALR 98
   : : || || | | | | | | | | | | | | | | | | | | | |
Db 99 YSRALLEISGFANQAVIANMAKEDRIADRLS-----HASLEEAS 141
   : : || || | | | | | | | | | | | | | | | | | | | |

Qy 99 LPPTSLKPDTRD-----QLQELCIPDLVGDLSAVWF---GSQRPL--LDSVAQQOGA 147
   : : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 142 LSPSOLRRFAHNDVTHLARLLASPCPGQOMV--VTEGVFSMDGSAPLAEIQVTTQHG 199
   : : | : | : | : | : | : | : | : | : | : | : | : | : |

Qy 148 WLPVADFRRVDVAITSALA-----RSLQPSVLM----- 179
   || | | | | | | | | | | | | | | | | | | | | | | | |
Db 200 WL-----MVDHAGTGVIGEGRGSCWLOKVRPELLVITFGFGVSGVAIVLCSSTV 251
   || | | | | | | | | | | | | | | | | | | | | | | | |

Qy 179 ---OLKSDGSAYRFEVPTAKFQELYSVALVKEMADLEK 216
   : : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 252 ADYLQFARHLIVSTSMPPAQALRASLAVIRSDSGDARR 292
   : : | : | : | : | : | : | : | : | : | : | : | : | : |

RESULT 8
US-08-875-062-3
; Sequence 3, Application US/08875062B
; Patent No. 5942431
; GENERAL INFORMATION:
; APPLICANT: YONEDA, TADASHI
; APPLICANT: HARUM, TAKADA
; APPLICANT: KEL, OHNO
; APPLICANT: JUNJI, SASUGA
; TITLE OF INVENTION: NOVEL LIPASE GENE AND PROCESS FOR THE

```


;; TITLE OF INVENTION: PRODUCTION OF LIPASE WITH THE USE OF THE SAME
;; FILE REFERENCE: 5059.204
;; CURRENT APPLICATION NUMBER: US/08/875,062B
;; CURRENT FILING DATE: 1997-07-10
;; EARLIER APPLICATION NUMBER: PCT/JP96/00426
;; EARLIER FILING DATE: 1996-02-23
;; NUMBER OF SEQ ID NOS: 16
;; SOFTWARE: FastSeq for Windows Version 3.0
;; SEQ ID NO 3
;; LENGTH: 335
;; TYPE: PRT
;; ORGANISM: Pseudomonas sp.
US-08-875-062-3

Query Match 7.2%; Score 81; DB 2; Length 335;
Best Local Similarity 24.9%; Pred. No. 0.23;
Matches 50; Conservative 32; Mismatches 85; Indels 34; Gaps 80
QY 21 RVSFGLAQLPPEVAMARLLGDLDRSTFRKLLKLVYSSLOGEDCRDGYQRLGVSAANLPEE 80
DB 168 RASLFNAQ---EHAFFASEEYVNOFTLERLAILHDPQLDQDKAERLERL--REGLPDE 222
QY 81 QLGLACGMH-TLLQALRLPPTSLKPDFTFDQQLQELCIPQDLVGLDASVVGSGORPLLD 139
DB 223 LQQLLVPLHQLTLLRQOTQQLLEQGAEP-----EQLRQLRL--NLVGPQAT-----ERLE 269
QY 140 SVAQOQGAWLPHVADFRWRVDVAISTALSRLPSVLMQLKSLDGSAYRFEVPTAKFOE 199
DB 270 ALDRQSEMDQLSGFNERNQRAISQPLGLADSKQAIEALLHFOFSEH-----ER 320
QY 200 LRYSVLVKEMADLEKRCR 220
DB 321 LRVS-----SLGLDSRAER 335

RESULT 9
US-08-530-950-8
; Sequence 8, Application US/08530950
; Patent No. 5736381
; GENERAL INFORMATION:
; APPLICANT: Davis, Roger J.
; APPLICANT: Raigneaud, Joel
; APPLICANT: Gupta, Shashi
; APPLICANT: Derijard, Benoit
; TITLE OF INVENTION: CYTOKINE-, STRESS-, AND
; TITLE OF INVENTION: ONCOPROTEIN-ACTIVATED HUMAN PROTEIN KINASE
; NUMBER OF SEQUENCES: 16
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson P.C.
; STREET: 225 Franklin Street
; CITY: Boston
; STATE: MA
; COUNTRY: USA
; ZIP: 02110-2804
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/530,950
; FILING DATE: 19-SEP-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Fasse, J. Peter
; REGISTRATION NUMBER: 32,983
; REFERENCE/DOCKET NUMBER: 07917/010001
; TELEPHONE: 617/542-5070
; TELEFAX: 617/542-8906
; TELEX: 200154

;; INFORMATION FOR SEQ ID NO: 8:
;; SEQUENCE CHARACTERISTICS:
; LENGTH: 393 amino acids
; TYPE: amino acid
; STRANDEDNESS: not relevant
; TOPOLOGY: linear
US-08-530-950-8

Query Match 7.0%; Score 79; DB 1; Length 393;
Best Local Similarity 20.0%; Pred. No. 0.5;
Matches 59; Conservative 42; Mismatches 88; Indels 106; Gaps 14;
QY 3 AVGAATPYLHHPGDSHSGRYVFLGAQLPPEVAMARLLGDLDRSTFRKLLKLVYSSLOGE 62
DB 106 AYGSVNRWVHKP-----SG-----QIMAVKIRSTVDEKEQKQLLDLVVMRSS 150
QY 63 DC-----RDG-----VQRLGVSA-----LPEOLGAL-LAGMHTL 92
DB 151 DCPYIVQYFALFREGDCWICMELMSTSPDKFYKYVYVLDVPIPEILKRLTATVKAL 210
QY 93 --LQALRLPPTSLKPDFTF--RDQLQELCIPQDLVGLDASVVGSGORPLLDVSAQOQGA 147
DB 211 MHLKENLKIHRDIKPSNILLDRSGMIKLC-----DFGISGLVDSIAKTRDA 258
QY 148 ----WLPHVADFR-----WRVDVAISTALSRLPSVLMQL-----SVLMQL----- 181
DB 259 GCRPYVAPERIDFSASRQGYVRSVMSGLITLYELATGRFPYKPNVSVFDQLTVVVKGD 318
QY 181 --KLSDGSAYRF-----EYPTAKFOELRYSVLVKEMADLEKRC 218
DB 319 PPQLNSSEEREFSPIFNVLNCLTKDESKRPYKLLKHPFILMYEERAVEVAC 373

RESULT 10
US-08-606-888A-5
; Sequence 5, Application US/08606888A
; Patent No. 5766913
; GENERAL INFORMATION:
; APPLICANT: Lin, Shuen-Fuh
; APPLICANT: Chlou, Chien-Ming
; APPLICANT: Chuang, Kuang-Hsiang
; TITLE OF INVENTION: CLONING, EXPRESSION AND NUCLEOTIDE
; TITLE OF INVENTION: SEQUENCE OF A AKALINE GENE FROM PSEUDOMONAS
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson P.C.
; STREET: 225 Franklin Street
; CITY: Boston
; STATE: MA
; COUNTRY: USA
; ZIP: 02110-2804
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/606,888A
; FILING DATE: 26-FEB-1996
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Tsao, Rocky Y.
; REGISTRATION NUMBER: 34,053
; REFERENCE/DOCKET NUMBER: 06840/003001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617/542-5070
; TELEFAX: 617/542-8906
; TELEX: 200154
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 340 amino acids

;; FILING DATE: April 10, 1992
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: US 07/745,206
;; FILING DATE: 15-AUG-1991
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: US 07/620,250
;; FILING DATE: 30-NOV-1990
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: US 07/482,384
;; FILING DATE: 20-FEB-1990
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: US 07/603,751
;; FILING DATE: 04-APR-1989
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: WO PCT/US89/01408
;; FILING DATE: 04-APR-1989
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: US 07/176,899
;; FILING DATE: 04-APR-1988
;; ATTORNEY/AGENT INFORMATION:
;; NAME: Seidman, Stephanie L.
;; REGISTRATION NUMBER: 33,779
;; REFERENCE/DOCKET NUMBER: 52516 (P519739)
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: (619)238-0999
;; TELEFAX: (619)238-0062
;; INFORMATION FOR SEQ ID NO: 53:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 1103 amino acids
;; TYPE: amino acid
;; STRANDEDNESS: single
;; TOPOLOGY: linear
;; MOLECULE TYPE: protein
;; FRAGMENT TYPE: internal
US-08-223-305C-53

Query Match 6.3%; Score 71.5; DB 2; Length 1103;
Best Local Similarity 23.1%; Pred. No. 18;
Matches 31; Conservative 19; Mismatches 59; Indels 25; Gaps 5;
QY 9 PYLHPGDSHGRVSYFGLAQLPPEVAMAR---LLGLDRSTRFKLK----- 54
Db 546 PNIONPKSQEPVTLDFDLAELENDIKVEIRKNMIDSEKTEFTLVKSODERYIDKGNR 605
QY 54 -FVSSVLOGEDCRGVQRLGVSNLANPEEQIGALLAGHHTLLOQALRLPPTSLKPDFTFRQ 112
Db 606 TYTWTVPVNGTD-----YSLALVLTYSFYIKAKLEETITQA-RYSET-LKPDNFEES 656
QY 113 LQELCIPQDLVGLD 126
Db 657 GYTFIAPRDYCNLD 670

RESULT 13
US-08-369-822C-27
; Sequence 27, Application US/08369822C
; Patent No. 6015660
; GENERAL INFORMATION:
; APPLICANT: Lipkin, W. I.
; APPLICANT: Briese, Thomas
; APPLICANT: Kliche, Stefanie
; APPLICANT: Schneider, Patrick A.
; APPLICANT: Stitz, Lothar
; APPLICANT: Schneemann, Anette
; TITLE OF INVENTION: Borna Disease Viral Sequences,
; TITLE OF INVENTION: Diagnostics and Therapeutics for Central Nervous
; TITLE OF INVENTION: System Diseases
; NUMBER OF SEQUENCES: 46
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fulbright & Jaworski, L.L.P.
; STREET: 865 South Figueroa Street, 29th Floor
; CITY: Los Angeles

;; STATE: California
;; COUNTRY: USA
;; ZIP: 90017-2571
;; COMPUTER READABLE FORM:
;; MEDIUM TYPE: Floppy disk
;; COMPUTER: IBM PC compatible
;; OPERATING SYSTEM: WINDOWS NT-WORDPERFECT 8.0
;; SOFTWARE: ASCII (DOS) TEXT
;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: US/08/369,822C
;; FILING DATE: 06-JAN-1995
;; CLASSIFICATION: 435
;; ATTORNEY/AGENT INFORMATION:
;; NAME: Churchill, Margaret A. (Ph.D.)
;; REGISTRATION NUMBER: 39,944
;; REFERENCE/DOCKET NUMBER: 1279-194XX
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: 213/892-9200
;; TELEFAX: 213/680-4518
;; INFORMATION FOR SEQ ID NO: 27:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 489 amino acids
;; TYPE: amino acid
;; TOPOLOGY: linear
;; MOLECULE TYPE: peptide
;; HYPOTHETICAL: NO
;; ANTI-SENSE: NO
US-08-369-822C-27

Query Match 6.2%; Score 70.5; DB 3; Length 489;
Best Local Similarity 26.0%; Pred. No. 6.9;
Matches 34; Conservative 15; Mismatches 49; Indels 33; Gaps 6;
QY 13 HPGDSHGRVSYFGLAQLPPEVAMARLLGLDRSTRFKLKLVVSSVLOGEDCRGVQRL 71
Db 349 HSGFYFYGKQYLVGLVQSLKTMAR-CGPLSDSIF-----DDLQGLASI 394
QY 72 GYSANLPEEQIGAL-----LAGMHTLL-----QALRLP-----PTSUKPDFTFRQ 113
Db 395 GVSFSGTSETRHPPSRWIASFHSMLAINLNQHLGFLPLGNIDISCFKKPLTFSEKL 454
QY 114 QELCIPQDLVGL 124
Db 455 IALITPQVLGG 465

RESULT 14
US-08-764-100-27
; Sequence 27, Application US/08764100
; Patent No. 5773700
; GENERAL INFORMATION:
; APPLICANT: van Grinsven J., Martinus Q.
; APPLICANT: De Haan, Petrus T.
; APPLICANT: Gielen L., Johannes J.
; APPLICANT: Peters, Dirk
; APPLICANT: Goldbach, Robert W.
; TITLE OF INVENTION: Improvements in or Relating to Organic
; NUMBER OF SEQUENCES: 27
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sandoz Agro, Inc
; STREET: 975 California Avenue
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:

```

; APPLICATION NUMBER: US/08/764,100
; FILING DATE: 06-DEC-1996
; CLASSIFICATION: 800
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/214,064
; FILING DATE:
; APPLICATION NUMBER: US/08/032,235
; FILING DATE: 17-MAR-1993
; APPLICATION NUMBER: GB 9206016.9
; FILING DATE: 19-MAR-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: No. 57737001s, Allen E.
; REGISTRATION NUMBER: 34,490
; REFERENCE/DOCKET NUMBER: 137-1061
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 354-3592
; TELEFAX: (415) 857-1125
; INFORMATION FOR SEQ ID NO: 27:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3218 amino acids
; TYPE: amino acid
; STRANDEDNESS: unknown
; TOPOLOGY: unknown
; MOLECULE TYPE: protein
; US-08-764-100-27

Query Match 6.2%; Score 70.5; DB 1; Length 3218;
Best Local Similarity 22.1%; Pred. No. 1.2e+02;
Matches 58; Conservative 41; Mismatches 94; Indels 69; Gaps 13;

QY 7 ATPYLHPGDSHGVSFLGAOLPPEVA-----AMA-RLLGDLDRSTFRKLLAFV 55
Db 2839 ASPLKSSERTYSERVA-LGLALAEVALASPSERVALALAVALLEGPRGLASNILEIL 2897

QY 56 VSS--LQGEDCDRGVORLGVSNLPEEQIGALLAGMHTLLQOALRLPTSLKPDFTFRDL 113
Db 2898 EASPGLNASNASPLNHSALAHII--SGLGLGLGLINT-----YRASNSERASPTHRSER 2950

QY 114 QELCIPQDLVGD-----ASVFGSGORP-----LLDSVAQO-----QGAWL 149
Db 2951 VALTRPSERPHETRPASPTVRYVALLYSSERPRPHEASNPHEILEALASERHISPEGLYS 3010

QY 150 -----PHVADFRRVDRVAISTSALARSLOPSVLMQLKSDGSAYRFEVPTAKFQLRYSV 204
Db 3011 ERPEHPERASPTHRV-----ALARGVALVALLEILELEPHEVALP-----HE 3054

QY 205 ALVLKEMADLEKRC-----ERRLQ 223
Db 3055 ALALEALATVRLCYSSERILE 3076

RESULT 15
US-08-435-675B-5
; Sequence 5, Application US/08435675B
; Patent No. 5710250
; GENERAL INFORMATION:
; APPLICANT: Ellis, Steven Bradley
; APPLICANT: Williams, Mark E.
; APPLICANT: Harpold, Michael Miller
; APPLICANT: Schwartz, Arnold
; APPLICANT: Brenner, Robert
; TITLE OF INVENTION: CALCIUM CHANNEL COMPOSITIONS AND METHODS
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Brown, Martin, Haller & McClain
; STREET: 1660 Union Street
; CITY: San Diego
; STATE: CA
; COUNTRY: USA
; ZIP: 92101-2926
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq Version 1.5
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/435,675B
; FILING DATE: 05-MAY-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/314,083
; FILING DATE: 28-SEP-1994
; APPLICATION NUMBER: US 07/914,231
; FILING DATE: 13-JUL-1992
; APPLICATION NUMBER: US 07/603,751
; FILING DATE: 08-NOV-1990
; ATTORNEY/AGENT INFORMATION:
; NAME: Seidman, Stephanie L.
; REGISTRATION NUMBER: 33,779
; REFERENCE/DOCKET NUMBER: 6362-53193
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 619-238-0999
; TELEFAX: 619-238-0062
; TELEX:
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1106 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; FRAGMENT TYPE: internal
; US-08-435-675B-5

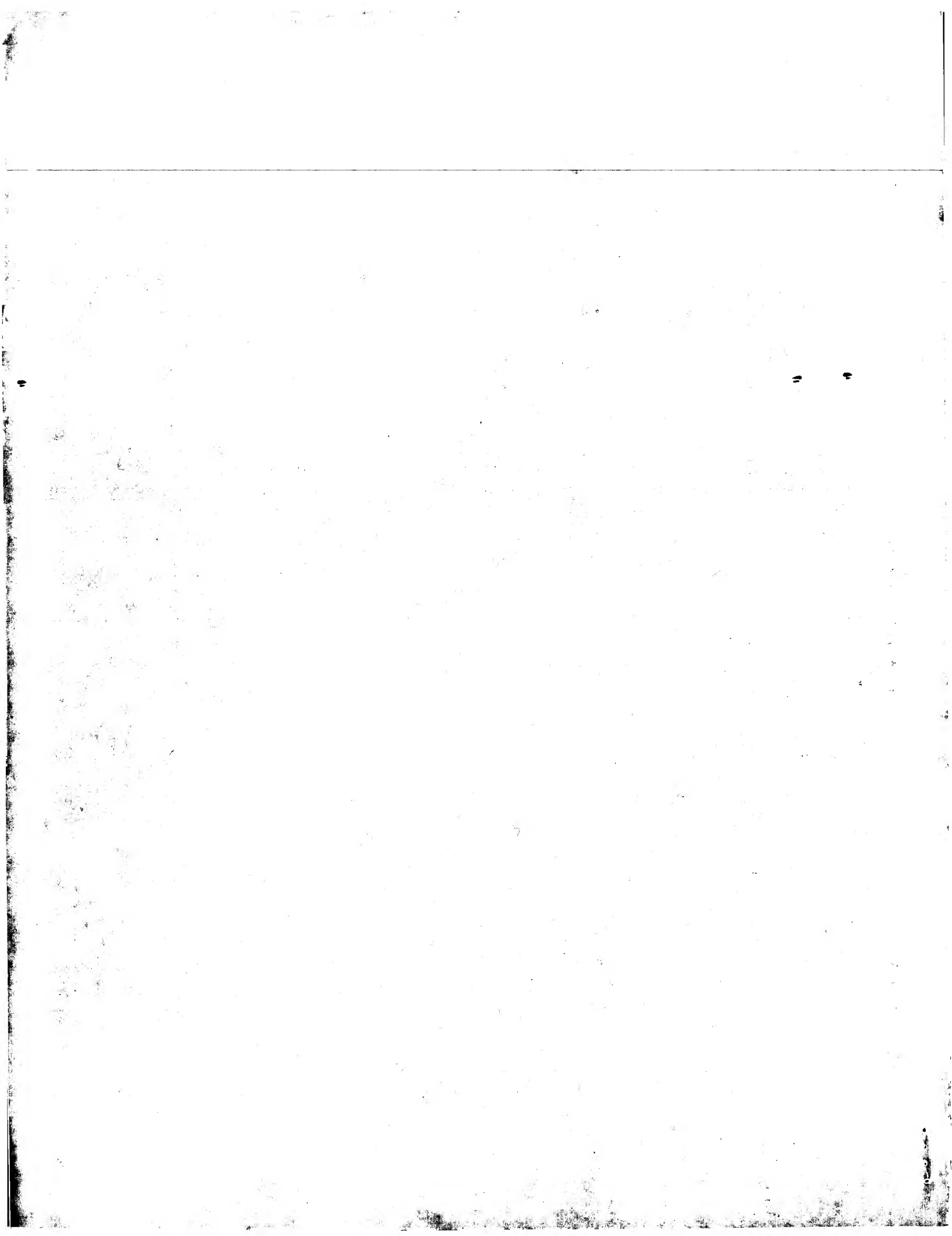
Query Match 6.2%; Score 70; DB 1; Length 1106;
Best Local Similarity 23.7%; Pred. No. 28;
Matches 32; Conservative 18; Mismatches 59; Indels 26; Gaps 5;

QY 9 PYLHHPGDSHGVSFLGAOLPPEVAAMAR---LLGLDRSTFRKLLKFFVSSLOGEDCDR 65
Db 548 PNVQNPXSQEPVTLDFLDALENDIKVEIRNMIDSGSEKFTRTLK-----SQDER 600

QY 66 -----DGVORLGVSNLPEEQIGALLAGMHTLLQOALRLPTSLKPDFTFRD 111
Db 601 YIDKGNRTYTWTPVNGTDYSSIALVLPTYSFYIKAKIETITQA-RYSET-LKPDNFE 658

QY 112 QLOELCIPQDLVGD 126
Db 659 SGYTFAPRDYCSDL 673

Search completed: May 15, 2000, 03:22:32
Job time: 5778 sec
```

F:604-675/Domain: annexin repeat homology <AX4>	
submitted to the EMBL Data Library, August 1995	
A:Reference number: 221762	
A:Accession: T38016	
A:Status: Preliminary; translated from GB/EMBL/DBJ	
A:Molecule type: DNA	
A:Residues: 1-863 <OLI>	
A:Cross-references: EMBL:298532; PIDN:CAB11082.1; GSPDB:GNO0066; SPDB:SPAC1B1.03c	
A:Experimental source: strain 972h-; cosmid c1B1	
C:Genetics:	
A:Gene: SPDB:SPAC1B1.03c	
A:Map position: 1	
Query Match 7.6%; Score 86; DB 2; Length 863;	
Best Local Similarity 20.9%; Pred. No. 13;	
Matches 42; Conservative 36; Mismatches 73; Indels 50; Gaps 6;	
QY 33	VAMARLLGDLDRSTRKLLKFFVSSLOGEDCR-----DGQRLGVSANLPEQLGALLA 87
Db	580 LTSIIRFGPDIFTSDDQIMNLLQTMOTAPKOSVVHEDVLLAIGAMNSLEQEVEYVP 639
QY 88	GMHTLLQALRLPPTSLKPDTRDQLQELC-IPQDLVGDLSVVFSGRSRPLLDLSVAQOQG 146
Db	640 SFVFFLSALS-----NEQEVQLCSVAVGLVDLARAL-----NA 674
QY 147	ANLPHVADFRWRVDVAISTALSRLQPSVL---MQLKSDGSAYRFEV----- 193
Db	675 KILPYCDEFMTRLVQDILQSSVLDRNYPAILSCFSDIALAIGAFAQTYLEAVNVLLQOAS 734
QY 193	-----PTAKFQELRYSLVALYL 208
Db	735 SVQAPPGANFSMIDYVDALRL 755
RESULT 5	
T14320	
calineurin inhibitor cain - rat	
C:Species: Rattus norvegicus (Norway rat)	
C:Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 20-Sep-1999	
C:Accession: T14320	
R:Lai, M.M.; Burnett, P.E.; Wolosker, H.; Blackshaw, S.; Snyder, S.H.	
J. Biol. Chem. 273, 18325-18331, 1998	
A:Title: Cain, a novel physiologic protein inhibitor of calcineurin.	
A:Reference number: Z17979; MUID:98325042	
A:Accession: T14320	
A:Status: preliminary; translated from GB/EMBL/DBJ	
A:Molecule type: mRNA	
A:Residues: 1-2182 <LAI>	
A:Cross-references: EMBL:AF061947; NID:g3323606; PID:g3323607; PIDN:AAC40176.1	
Query Match 7.6%; Score 85.5; DB 2; Length 2182;	
Best Local Similarity 20.8%; Pred. No. 47;	
Matches 49; Conservative 34; Mismatches 90; Indels 63; Gaps 8;	
QY 10	YLHPGDSHSGRVSLGAQLPPEV-----AAMARLLGDLDRSTRFK-LLKFEV 56
Db	1236 YLHEAARYPKIHY---HNPELAMEALEVYFRLHASILKLLGCKPDGVSVAELVSEFMK 1292
QY 57	SSLOG-----EDCRDGVQRL-GVSANLPEQLGALLAGMHTL 92
Db	1293 EAAEGPFARGEKNTPKASEKEKACLVDESHSSAGTLPGPCASLPSSSGPGLTSPPYTA 1352
QY 93	L-----QQALRLPPTSLKPDTRDQLQELCIPDQLVGDLSVVFSGRSRPL- 138
Db	1353 TPIDHDYVKCKKPRQQA---TPDDRSQDSTAVLSDSSSTQDFFNEPTSLIDGSKLLPE 1409
QY 138	-----LDSVAQOQGAWLPHVADFRWRVDVAISTALSRLQPSVLQMLKSDGSAY 188
Db	1410 KRISGLSAQPSGKDLPCGPTEERGKTEESLESTFAFRWEPVSKPVPADSSASAY 1465
RESULT 6	

Query Match 7.2%; Score 82; DB 2; Length 399;
Best Local Similarity 20.2%; Pred. No. 11;
Matches 59; Conservative 44; Mismatches 89; Indels 100; Gaps 13;

7-oxononanoate synthase (EC 2.3.1.47) - *Escherichia coli*
 Alternate names: 7-KAP synthetase; 7-keto-8-amino pelargonic acid synthetase
 Species: *Escherichia coli*
 Date: 30-Jun-1990 #sequence_revision 30-Jun-1990 #text_change 18-Jun-1999
 Accession: D32025; H64813
 Otsuka, A.J.; Buoncrisistiani, M.R.; Howard, P.K.; Flamm, J.; Johnson, C.; Yamamoto, Y.
 Biol. Chem. 263, 19577-19585, 1988
 Title: The *Escherichia coli* biotin biosynthetic enzyme sequences predicted from the
 Reference number: A32025; MUID:89066784

	Query Match	7.28;	Score 81.5;	DB 2;	Length 800;	
	Best Local Similarity	19.4%;	Pred. No. 30;	Mismatches 68;	Indels 97;	Gaps 10;
	Matches 49;	Conservative				
Qy	32	EYAAARLLGLDGRS-----TRKLLKFVSSSLAGEDC-----RCGVQLGSANLPE	79			
Ddb	55	ESAAFLFVLQDMQESNMERNVVFKILRRLCTILHADRCSLFWYQRNRGAEI-----	107			
Qy	80	EOLGALLAGMHTLLQOALRLPPTLSLKPTDFRDQQLCIPQDLVGDSLAVVFSGSORPLLD	139			
Ddb	107	-----ATRL--FSVPQDSL---LBDCLVPPD-----SEIVFPDLDIGTV	140			
Qy	140	SVAQ-----OOGAWLPHVADFRWRVDVAISTALSRLQ-----PSVLMLQLSDGS-	187			
Ddb	141	HVAQTKKMINVDVAECPFHSSFADFELTDYVTKNLTSTPMNGKDVAIVAMVKNLDDGPC	200			
Qy	187	-----AYREFVPTAFQELRYSVVALVLKEMADLEKR-----	218			
Qy	203	CPCACDECHETXVI NEATNI NKIVTH SYLNCTRTRRGVOILNSANKVFBELDTIEROFHKA	260			

DB 261 FTVRAYLNCR 272

RESULT 11

S30762

3', 5'-cyclic-GMP phosphodiesterase (EC 3.1.4.35) beta chain - mouse

C; Species: Mus musculus (house mouse)

C; Date: 22-Nov-1993 #sequence_revision 19-Oct-1995 #text_change 20-Aug-1999

C; Caccession: S30762; S13031; S13121

R; Baehr, W. submitted to the EMBL Data Library, June 1991

A; Reference number: S30762

A; Accession: S30762

A; Status: preliminary

A; Molecule type: mRNA

A; Residues: 1-856 <BAE>

A; Cross-references: EMBL:X60133; NID:g53595; PIDN:CAA42719.1; PID:g53596

R; Baehr, W.; Champagne, M.S.; Lee, A.K.; Pittler, S.J. FEBS Lett. 278, 107-114, 1991

A; Title: Complete cDNA sequences of mouse rod photoreceptor cGMP phosphodiesterase al-
pha of the beta-subunit gene.

A; Reference number: S13030; S13031; S130581

A; Accession: S13030; S13031; S130581

A; Residues: 1-559, 'G', 561-856 (Ba2>
A; Cross-references: EMBL: X60133
R; Bowes, C.; Li, T., 1990
Nature 347, 677-680. 1990
A; Title: Retinal degeneration in the rd mouse is caused by a defect in the beta subun
A; Reference number: S13121; MUID: 91015387
A; Accession: S13121
A; Molecule type: mRNA
A; Residues: 1-2, 'X', 4, 'G', 6-18, 'S', 20-48, 'DV', 51-157, 'T', 159-175, 'C', 177-231, 'R', 233-
A; Cross-references: EMBL: X55968; NID: G53616; PIDN: CAA39439.1; PID: G53617
A; Note: The authors translated the codon AGA for residue 232 as Glu
C; Superfamily: 3', 5'-cyclic-GMP phosphodiesterase alpha chain; 3', 5'-cyclic-nucleotid
C; Keywords: cGMP binding; phosphoric diester hydrolase

Query Match	7.2%;	Score 81.5;	DB 2;	Length 856;
Best Local Similarity	19.4%;	Pred. No. 32;		

[illegible]

Db 107 -----ATRL--FSVQPSL--LEDCLVPPD-----SEIVFPLDIGIVG 140
QY 140 SVAQ-----OQAWLPHVADFRWDVAISTSAARSIQ-----PSVLMOLKLSGDS- 187
Db 141 HVAQTKMINVDVAECPPHSSFADELTDVTKNIIUSTPLMNGKDVAVIMAVNKLDGPGC 200
QY 187 -----AYRFEVPTAKFQELRYISVALVKEMADLEKR----- 218
Db 201 FTSEDEVDFTKYLNFATLNKLIHYLSLHNCETRRGQVLLWSANKVFEELTDIERQFHKA 260
QY 218 -----CER 220
Db 261 FYTVRAYLNGER 272

RESULT 12
A53824
nuclear pore membrane protein POM152 - yeast (Saccharomyces cerevisiae)
N:Alternate names: protein YM9553.05; protein YMR129w
C:Species: Saccharomyces cerevisiae
C>Date: 07-Oct-1994 #sequence_revision 07-Oct-1994 #text_change 24-Nov-1999
C:Accession: A53824; B53824; S53059
R:Wozniak, R.W.; Blobel, G.; Rout, M.P.
J. Cell Biol. 125, 31-42, 1994
A:Title: POM152 is an integral protein of the pore membrane domain of the yeast nuclear
A:Reference number: A53824; MUID:94186543
A:Accession: A53824
A:Molecule type: DNA
A:Residues: 1-1337 <WQ>
A:Cross-references: GB:Z31592; NID:9473153; PID:9473154
A:Accession: B53824
A:Molecule type: protein
R:Badcock, K.; Churcher, C.
submitted to the EMBL Data Library, March 1995
A:Reference number: S53055
A:Accession: S53059
A:Molecule type: DNA
A:Residues: 1-1337 <BAD>
A:Cross-references: EMBL:Z48622; NID:g728663; PID:g728668; MIPS:YMR129w
C:Genetics:
A:Gene: SGD:POM152
A:Cross-references: SGD:S0004736; MIPS:YMR129w
A:Map position: 13R
C:Superfamily: Saccharomyces cerevisiae nuclear pore membrane protein POM152
C:Keywords: blocked amino end; duplication; glycoprotein; nuclear membrane; transmembran
F:176-195/Domain: transmembrane #status predicted <TM>
F:180/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 7.2%; Score 81.5; DB 2; Length 1337;
Best Local Similarity 25.9%; Pred. No. 56;
Matches 45; Conservative 28; Mismatches 72; Indels 29; Gaps 9;
QY 56 VSSIQGDCRGVQRLGVGSANLPEEQIGA---LLAGMHTLLOQALRLPPTSLKPDTRDQ 112
Db 380 IASITGTGSND--RCIGSDNVSEIQGVPPMKLAYSKIIVGQTFSSVSDSLQPEYFESP 437
QY 113 LQELICPDQLV-GDLASVFGSQRPPL---LDSVAQQGAWLPHVADFRWRVDVAISTSAL 168
Db 438 LQSSKSKQSFQGLNDLKNQPNVINDSSITDQ-----KPAYKIDKI--TDGL 488
QY 169 ARSIQ-PSVLMOLKLSGDSAYRF---EVPTAKFOE-----LRYISVALVKEM 211
Db 489 GNVVDFTSLPEELKKRYDLSYFNFNHVEVPRAALEERDPKSPKRSATVFEI 542

RESULT 13
E69440
conserved hypothetical protein AF1526 - Archaeoglobus fulgidus
C:Species: Archaeoglobus fulgidus
C>Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 29-Sep-1999

C:Accession: E69440
R:Klenk, H.P.; Clayton, R.A.; Tomb, J.F.; White, O.; Nelson, K.E.; Ketchum, K.A.; Dod
.: Fleischmann, R.D.; Quackenbush, J.; Lee, N.H.; Sutton, G.G.; Gill, S.; Kirkness, E
Glodek, A.; Zhou, L.; Overbeek, R.; Gocayne, J.D.; Weidman, J.F.; McDonald, L.
Nature 390, 364-370, 1997
A:Authors: Overbeek, T.; Cotton, M.D.; Spriggs, T.; Artach, P.; Kaine, B.P.; Sykes,
Smith, H.O.; Woese, C.R.; Venter, J.C.
A:Title: The complete genome sequence of the hyperthermophilic, sulfate-reducing arch
A:Reference number: A69250; MUID:98049343
A:Accession: E69440
A>Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-145 <KLE>
A:Cross-references: GB:AE000997; GB:AE000782; NID:g2689320; PIDN:AA89722.1; PID:g264
C:Superfamily: hypothetical protein AF1526

Query Match 7.2%; Score 81; DB 2; Length 145;
Best Local Similarity 30.7%; Pred. No. 3.9;
Matches 35; Conservative 21; Mismatches 40; Indels 18; Gaps 5;
QY 36 MARLLGDLDRS-----TRKLLKLVVSSIQGDCRGVQRLGVGSANLPEEQALGAGMH 90
Db 1 MERILLVDDTGTGEIAFOKLAKLAEDGLRGVEYIYIREMEVPPFVPEEK---ELAAVH 57
QY 91 TLLQALRLPPTSLKPDTRDQLQELICPDQLVGLASVFGSQRPPLDLSVAQQ 144
Db 58 RLMTQSMK-----KLEGFKNOLERAGLEK---VSDV-SVVFCKYADRLLLVEKQ 101

RESULT 14
S61303
heat shock protein 63 - Neisseria meningitidis (fragment)
C:Species: Neisseria meningitidis
C>Date: 27-Apr-1996 #sequence_revision 17-May-1996 #text_change 26-Aug-1999
C:Accession: S61303; S40249
R:Pannekoek, Y.; Dankert, J.; van Putten, J.P.M.
Mol. Microbiol. 15, 277-285, 1995
A:Title: Construction of recombinant neisserial Hsp60 proteins and mapping of antigen
A:Reference number: S61300; MUID:95264914
A:Accession: S61303
A>Status: nucleic acid sequence not shown
A:Molecule type: DNA
A:Residues: 1-544 <PAN>
A:Cross-references: EMBL:Z22956; NID:g438205; PIDN:CAA80532.1; PID:g438206
C:Superfamily: chaperonin groEL
C:Keywords: heat shock; stress-induced protein

Query Match 7.2%; Score 81; DB 2; Length 544;
Best Local Similarity 19.9%; Pred. No. 20;
Matches 61; Conservative 45; Mismatches 84; Indels 116; Gaps 14;
QY 25 LGAQLPPEVAANAR-LLGD-----LDRSTFRKLLKLVVSSIQGDCRGVQRL----- 72
Db 69 MGAQMYKEVASKINDVAGDGTATVLAQSIIVAGMKYVTAGNPNPTDLKRGIDKAVAAVL 128
QY 72 -----GVSANLPEQALGALLA-----GMHTL----- 93
Db 129 ELKNTAKPCDTSKETAQVGSISAN-SDEQVGAITAEAMEKVGKEGVIIVDGKSLNEL 187
QY 93 -LQOALRLPPTSLKP-----DTRDLOQE----- 116
Db 188 DVVEGQDFRGYLSPIFFINDAEKQIAGLNPVLLPEKKISNIRDLPLVLEKVAKASRPL 247
QY 116 LCTPQDLVGD-LASVFGSQRPPLDLSVAQQGAWLPHVADFRWRV--DVAISTSALARS 172
Db 248 LIIAENVEGALATLVNNIRGILKTVAVK---APFGDRRKAMLDQIALLTGGTVISE 303
QY 173 OPSVLMLKLSD--GSAYRFEVP-----TAKFOELRYISVALVKEMADL 214
Db 304 EVGLFLEKATLDNLGQAKRIEIGKENTTIDGFGDAAQAEARFAEIRQQIETATSDY-DK 362

OY 215 EXRCER 220
 II II
 Db 363 EXLQER 368

RESULT 15

E72339
 Hypothetical protein - Thermotoga maritima (strain MSB8)
 C:Species: Thermotoga maritima
 C>Date: 11-Jun-1999 #sequence_revision 11-Jun-1999 #text_change 11-Jun-1999
 C:Accession: E72339
 R:Nelson, K.E.; Clayton, R.A.; Gill, S.R.; Gwinn, M.L.; Dodson, R.J.; Haft, D.H.; Hickey
 Garrett, M.M.; Stewart, A.M.; Cotton, M.D.; Pratt, M.S.; Phillips, C.A.; Richardson, D.;
 C.M.
 Nature 399, 323-329, 1999
 A:Title: Evidence for lateral gene transfer between Archaea and Bacteria from genome seq
 A:Reference number: A72200; MUID:99287316
 A:Accession: E72339
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-161 <ARN>
 A:Cross-references: GB:AE001744; GB:AE000512; NID:g4981254; PID:g4981263; TIGR:TM0738
 A:Experimental source: strain MSB8
 C:Genetics:
 A:Gene: TM0738

Query Match 7.1%; Score 80.5; DB 2; Length 161;
 Best Local Similarity 27.7%; Pred. No. 5;
 Matches 26; Conservative 20; Mismatches 35; Indels 13; Gaps 3;
 OY 1 MSVAGCAATP-YLHHPCDSHGSRVSEFGAQLPPEVAAMARLLGDLDRSTFRKLLKFWVSSL 59
 Db 69 LAAVAGRTQVFYITPTKSSERTHSLGSELQDKQOEIERKILELRQ-----LKTETDPL 123
 OY 60 QGEDCRDGVQRLGVSANL-----PEEQLGALL 86
 Db 124 ERERLREDLERLKLALNLIKASLRAPELLVGVL 157

Search completed: May 15, 2000, 03:23:38
 Job time: 5548 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

Protein - protein search, using sw model

Run on: May 15, 2000, 05:34:20 ; Search time 44.1 seconds
(without alignments)
154.692 Million cell updates/sec

Title: US-09-223-796-4
Perfect score: 1132
Sequence: 1 MSAVGAATPVLHPGDSHG.....ALVLKEMADLEKRCERLQD 224

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 83857 seqs, 30454973 residues

Total number of hits satisfying chosen parameters: 83857

Minimum DB seq length: 0
Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database : SwissProt_38.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	86	7.6	863	1	IMB1 SCHPO
2	85.5	7.6	757	1	YHGF NEIME
3	85.5	7.6	2182	1	CAB1 RAT
4	83	7.3	397	1	MPK4 MOUSE
5	82.5	7.3	964	1	UL70 MCMVS
6	82	7.2	399	1	MPK4 HUMAN
7	81.5	7.2	384	1	BIOF_ECOLI
8	81.5	7.2	1337	1	P152 YEAST
9	81	7.2	344	1	CH60 NEIME
10	80.5	7.1	446	1	PFES_PSEAE
11	80	7.1	278	1	BGLG_ECOLI
12	80	7.1	1947	1	MYSC_CABEL
13	79.5	7.0	890	1	GUND_ECOLI
14	79.5	7.0	918	1	QALS_NEUCR
15	79	7.0	853	1	CNRB_BOVIN
16	78	6.9	234	1	NOLW_RHIFR
17	78	6.9	234	1	NOLW_RHISN
18	78	6.9	1241	1	KPBL_MOUSE
19	77.5	6.8	3329	1	BRC2_MOUSE
20	77	6.8	195	1	INO2_HORSE
21	77	6.8	230	1	HEMX_HAETN
22	77	6.8	318	1	DHP2_MYCTU
23	77	6.8	373	1	LEU3_CANNA
24	77	6.8	544	1	CH60_NEIGO
25	76.5	6.8	302	1	VRO_BBP2
26	76.5	6.8	361	1	RFL_COXBU
27	76.5	6.8	377	1	YGBD_ECOLI
28	76.5	6.8	496	1	NIFE_RHISN
29	76.5	6.8	3137	1	CA36_CHICK
30	76.5	6.8	4568	1	DYHB_CHLRE
31	76	6.7	1239	1	V120_EBV
32	75.5	6.7	544	1	CH60_NEIFL
33	75.5	6.7	548	1	CH60_ACIPI
34	75.5	6.7	592	1	LAM2_MOUSE

35	75.5	6.7	1391	1	YLD5_CABEL
36	75	6.6	359	1	COBT_ECOLI
37	75	6.6	493	1	CSD_RAT
38	75	6.6	772	1	NFL1_HUMAN
39	75	6.6	1101	1	GUNC_CELFI
40	75	6.6	2594	1	7LES_DROVI
41	74.5	6.6	1189	1	HAIR_HUMAN
42	74	6.5	220	1	RP29_HUMAN
43	74	6.5	446	1	MPK2_XENLA
44	74	6.5	551	1	CH60_BUCAP
45	74	6.5	1301	1	DDX9_CABEL

ALIGNMENTS

RESULT 1
IMB1 SCHPO
ID IMB1 SCHPO STANDARD; PRT; 863 AA.
AC O13864;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DE IMPORTIN BETA-1 SUBUNIT (KARYOPHERIN BETA-1 SUBUNIT) (IMPORTIN 95)
GN SPAC1B1.03C.
OS Schizosaccharomyces pombe (Fission yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetales;
OC Schizosaccharomycetaceae; Schizosaccharomycetes.
RN (1)
RP SEQUENCE FROM N.A.
RC STRAIN-972;
RA Oliver K., Harris D., Barrell B.G., Rajandream M.A., Wood V.;
CC Submitted (AUG-1997) to the EMBL/GenBank/DBJ databases.
CC FUNCTION: REQUIRED FOR NUCLEAR PROTEIN IMPORT AND MEDIATES DOCKING
CC OF IMPORT SUBSTRATE TO DISTINCT NUCLEOPORINS. SERVES AS RECEPTOR
CC FOR NUCLEAR LOCALIZATION SIGNALS.
CC -!- SUBUNIT: FORMS A COMPLEX WITH IMPORTIN ALPHA SUBUNIT (BY
CC SIMILARITY).
CC -!- SUBCELLULAR LOCATION: CYTOPLASMIC AND NUCLEAR ENVELOPE (BY
CC SIMILARITY).
CC -!- SIMILARITY: BELONGS TO THE IMPORTIN BETA CHAIN FAMILY.
CC -!- SIMILARITY: CONTAINS 11 HEAT REPEATS.

This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation
at the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See [http://www.isb-sib.ch/](http://www.isb-sib.ch/announce/)
or send an email to license@isb-sib.ch).

EMBL; 298532; CAB11082.1;
Transport; Protein transport; Repeat; Nuclear protein.
KW DOMAIN 124 482 HEAT REPEATS DOMAIN 1.
FT DOMAIN 329 342 IAB DOMAIN.
FT DOMAIN 334 419 RAN-GTP BINDING (BY SIMILARITY).
FT DOMAIN 592 717 HEAT REPEATS DOMAIN 2.
FT DOMAIN 315 320 POLY-ALA
SQ SEQUENCE 863 AA; 94747 MW; 9502D72F5A1C6839 CRC64;

Query Match 7.6; Score 86; DB 1; Length 863;
Best Local Similarity 20.9%; Pred. No. 4.1;
Matches 42; Conservative 36; Mismatches 73; Indels 50; Gaps 6;
QY 33 VAAMARLLGDDLRSTRKLLMFVYSSSQGEDCR-----DGVRQLGVSNLPEEQGALLA 87
Db 580 LSIIRRGPDITSSQIMNLLQTQAPKQSVVHEDVLLAIGAMMNSLEEQEVYVP 639
QY 88 GMHTLQALRLPTSLKPTFFDQQLQELC-IPQILVGLDASVFGSORPLDSDVAQQG 146
Db 640 SFVPLFLSSALS-----NEQYQLCSVAVGLVGLDALAL-----NA 674

RA Zon L.I.;
RL Submitted (DEC-1996) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: DUAL SPECIFICITY KINASE THAT ACTIVATES THE JUN KINASES
CC MAPK1 (ERK2) AND MAPK9 (JNK2) AS WELL AS MAPK14 (P38) BUT NOT
CC -!- PTM: ACTIVATED BY PHOSPHORYLATION ON SER/THR BY MAP KINASE
CC KINASES (BY SIMILARITY).
CC -!- TISSUE SPECIFICITY: IS EXPRESSED ABUNDANTLY IN THE ADULT
CC BRAIN AND MUSCLE.
CC -!- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.
CC MAP KINASE KINASE SUBFAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC use by non-profit institutions. There are no restrictions on its
CC modified and this statement as long as its content is in no way
CC or send an email to license@isb-sib.ch).
CC EMBL: U18310; AAB1554.1;
CC HSP: P00518; LPHK
CC MGD: MGI:107199; SERK1.
CC PROSITE: PS00107; PROTEIN_KINASE_ATP; 1.
CC PROSITE: PS00108; PROTEIN_KINASE_ST; 1.
CC PFAM: PF00069; PKINASE; 1.
CC Transferase; Serine/threonine-protein kinase; Tyrosine-protein kinase;
CC ATP-binding; Phosphorylation.
CC DOMAIN 5 17 GLX/SER-RICH.
CC NP_BIND 100 366 PROTEIN KINASE.
CC BINDING 103 111 ATP (BY SIMILARITY).
CC ACT_SITE 129 127 ATP (BY SIMILARITY).
CC MUTAGEN 227 227 BY SIMILARITY.
CC K-:R: LOSS OF ATP-BINDING ACTIVITY.
CC SEQUENCE 397 AA; 44113 MW; B99C6688184E5B3D CRC64;

Query Match
Best Local Similarity 7.3%; Score 83; DB 1; Length 397;
Matches 59; Conservative 44; Mismatches 89; Indels 100; Gaps 13;
QY 3 AVGAATPYLHGDSGRVSFLGALPPEVAMARLLGDLDRSTFRKLLKFFVSSLOGE 62
Db 110 AYGSVNKKVHKP---SG-----QIMAVKRIIRSTVDEKEQQLMDLDVVMRGS 154
QY 63 DC-----RDG-----VORLGVSN-----LPEQIGAL-LAGMHTL 92
Db 155 DCPYIVQYGFREGDCWICMELSTSFDFYKYVSVLDDVPEEILGKITLATVKAL 214
QY 93 --LQOALRLPPTSLKPDFTDQLQELCIPQDLVGLASVFGSQRPPLDSVAOQOAWL- 150
Db 215 NMLKENLKIHRDKPSN-----ILLDRSGNTKLCDFGIGQLVDSIAKTRDAGR 265
QY 150 PHVADFR-----NRVDVAISTALSARSLOP-----SVLMQL-----K 181
Db 266 PYMAPERIDPSASRGYDVRSDVWSLGLITLDELATGRFPYKRWNSVFDQLTVQVGGDPQ 325
QY 182 LSDGSAYRF-----EYPTAKFQELRYSLVALVKEMADLEKRC 218
Db 326 LSNSEERFSPSFNFVNLCLTQDESKRKYKELLKHPFILMTEETVEVAC 377
RESULT 5
UL70_MCMVS
ID UL70_MCMVS
AC Q89153; STANDARD; PRT; 964 AA.
DT 15-JUL-1998 (Rel. 36, Created)
DE 15-JUL-1998 (Rel. 36, Last sequence update)
DE HELICASE/PRIMASE COMPLEX PROTEIN (PROBABLE DNA REPLICATION PROTEIN
GN UL70).

OS Murine cytomegalovirus (strain Smith).
OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
RN Betaherpesvirinae; Muromegalovirus.
RP SEQUENCE FROM N.A.
RX MEDLINE; 96082764.
RA Messerie M., Rapp M., Lucin P., Koszinowski U.H.;
RT "Characterization of a conserved gene block in the murine
RT cytomegalovirus genome";
CC Virus Genes 10:73-80(1995).
CC -!- FUNCTION: INVOLVED IN DNA REPLICATION (BY SIMILARITY).
CC -!- SIMILARITY: BELONGS TO FAMILY THAT GROUPS TOGETHER HSV-1 UL52,
CC EBV-1 7, EBV BSFL1, HVS-1 56, VZV 6, HCMV AND MCMV UL70.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC use by non-profit institutions. There are no restrictions on its
CC modified and this statement as long as its content is in no way
CC or send an email to license@isb-sib.ch).
CC EMBL: L07319; AAA96665.1;
CC Helicase; DNA replication.
CC SEQUENCE 964 AA; 109704 MW; 70605300E1D85864 CRC64;
Query Match
Best Local Similarity 7.3%; Score 82.5; DB 1; Length 964;
Matches 51; Conservative 21; Mismatches 67; Indels 47; Gaps 11;
QY 30 PPEVAA-----MARLLGDLDRSTF--RKLL-KFVVSLOGEDCRDQVRLGVSNALP--- 79
Db 641 PPTAAAGTQTKRLAGILDHTCLDRDLVCKLNAISHPGCECFDTGIYSHGRSINMPLY 700
QY 79 --EQIGALLAGMHTLLOQALRLPPTSLKPDFTDQ-----LQELCIPQDLVGLASVWF 131
Db 701 KLDASGLTL--HS-----RLNPIFIVPAGYDRPAEFVLOQLC-PQNL-----TH 743
QY 132 GSORPLDSVAOQOAWLPHVAD-----FRMYDVVAISTALSARSLOPSVLMQLK 181
Db 744 HGRPQRDGSADQLTEVLVHTDTRACADSDGNFLQSRARRAMSSGRL--PLGPLLRAHLS 801
QY 182 LSDGSA 187
Db 802 LESGQS 807
RESULT 6
MPK4_HUMAN
ID MPK4_HUMAN
AC P45985; STANDARD; PRT; 399 AA.
DT 01-NOV-1995 (Rel. 32, Created)
DE 01-NOV-1995 (Rel. 32, Last sequence update)
DE DUAL SPECIFICITY MITOGEN-ACTIVATED PROTEIN KINASE KINASE 4
DE (EC 2.7.1.1) (MAP KINASE KINASE 4) (JNK ACTIVATING KINASE 4
DE TERMINAL KINASE KINASE 1) (JNKK) (SAPK/JERK KINASE 1) (SEK1).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Primates; Catarrhini; Hominidae; Homo.
RP SEQUENCE FROM N.A.
RX MEDLINE; 95232504.
RA Lin A., Minden A., Martinetto H., Claret F.-X., Lange-Carter C.,
RT "Identification of a dual specificity kinase that activates the Jun
RL kinases and p38-Mpk2";
RN Science 268:286-290(1995).
RP SEQUENCE FROM N.A.
RC TISSUE-BRAIN,

MEDLINE; 95141073.
X A Derlward B., Raigneaud J., Barrett T., Wu I.-H., Han J.,
DR Ulevitch E.J., Davis R.J.;
DR "Independent human MAP-kinase signal transduction pathways defined by
CC MEK and MKK isoforms";
CC Science 267:682-685(1995).
CC [3]
SEQUENCE FROM N.A.
RR MEDLINE; 98283389;
RA Su G.H., Hilgers W., Shekher M.C., Tang D.J., Yeo C.-J., Hruban R.H.,
RA Kero S.E.;
RA "Alterations in pancreatic, biliary, and breast carcinomas support
RT MKKA as a genetically targeted tumor suppressor gene.";
RL Cancer res. 58:2339-2342(1998).
CC - !- FUNCTION: DUAL SPECIFICITY KINASE THAT ACTIVATES THE JUN KINASES
CC MAPK8 (JNK1) AND MAPK9 (JNK2) AS WELL AS MAPK14 (P38) BUT NOT
CC MAPK1 (ERK2) OR MAPK3 (ERK1).
CC - !- PTM: ACTIVATED BY PHOSPHORYLATION ON SER/THR BY MAP KINASE KINASE
CC KINASES (BY SIMILARITY)
CC - !- TISSUE SPECIFICITY: ABUNDANT EXPRESSION IS SEEN IN THE SKELETAL
CC MUSCLE. IT IS ALSO WIDELY EXPRESSED IN OTHER TISSUES.
CC - !- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.
CC MAP KINASE KINASE SUBFAMILY.

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation at the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/) or send an email to license@isb-sib.ch.)

CC CC EMBL; UI7743; AAC50127.1; --
CC DR EMBL; L36870; AAC41719.1; --
CC DR EMBL; AF070090; AAC24130.1; JOINED.
CC DR EMBL; AF070080; AAC24130.1; JOINED.
CC DR EMBL; AF070081; AAC24130.1; JOINED.
CC DR EMBL; AF070082; AAC24130.1; JOINED.
CC DR EMBL; AF070083; AAC24130.1; JOINED.
CC DR EMBL; AF070084; AAC24130.1; JOINED.
CC DR EMBL; AF070085; AAC24130.1; JOINED.
CC DR EMBL; AF070086; AAC24130.1; JOINED.
CC DR EMBL; AF070087; AAC24130.1; JOINED.
CC DR EMBL; AF070088; AAC24130.1; JOINED.
CC DR EMBL; AF070089; AAC24130.1; JOINED.
CC DR HSPP; FLI362; IFGK.
CC MIN; 601335; --
CC DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
CC DR PROSITE; PS00108; PROTEIN_KINASE_ST; 1.
CC DR PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
CC DR PFAM; PF00069; pkinase; 1
CC DR Transferase: Serine/threonine.
KW ATP-binding; phosphorylation.
KW DOMAIN 5
FT DOMAIN 102 367 GLY/SER-RICH.
FT NP_BIND 105 113 ATG (BY SIMILARITY).
FT BINDING 131 131 ATG (BY SIMILARITY).
FT ACT_SITE 229 229 BY SIMILARITY.
F-I ACCT_SEQ 399 AA; 44287 MW; A472537E2F26770B CRC64;

Page 1

[illegible]

QY	93	---	---LQALRLPPTSLKPTDFDQLQELCIQDPLDGLDASVVFQSGRPDLDSVACQQAQNL-	150
DB	217	NHLKENLKIIRHDIKPSN	-----ILLDRSGNIKLCDFGISQLVDSIAKTRDAGCR	267
QY	217	NHLKENLKIIRHDIKPSN	-----WVVDVAIGTSALARSLOP	---SVLMOL-----K 181
QY	150	PHVADSR	-----WVVDVAIGTSALARSLOP	---SVLMOL-----K 327
DB	268	PYAPAEIDPSASRGQYDVRSDVNSLGITIVELATGRFPYKPNVSDLTQVVKGDPPQ	218	
QY	182	LDGSAVRF	-----EYPTAKFOELRYSAVALVKEMADLEKRC	218
DB	328	LSNSEERFSPSINEVNLCTKDSRPRKYKELLKHPFILMYEERAVEVAC	379	

RESULT 7

BIOT_ECOLI STANDARD; PRT: 384 AA.

ID B10F_ECOLI

AC P12998:

DT 01-JAN-1990 (Rel. 13, Created)

DT 01-JAN-1990 (Rel. 13, Last sequence update)

DT 15-DEC-1998 (Rel. 37, Last annotation update)

DE 8-AMINO-7-OXONONANOATE SYNTHASE (EC 2.3.1.47) (7-KETO-8-AMINO-

DE PELARGONIC ACID SYNTHETASE) (7-KAP SYNTHETASE) (L-ALANINE -PIMELYL COA

DE LIGASE).

GN Escherichia coli.

OS Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae.

OC Escherichia.

OC [1]

RN SEQUENCE FROM N.A.

RP MEDLINE; 89066784.

RA Otsuka A., Buonocristiani M.R., Howard P.K., Flamm J., Johnson O.,

RA Yamamoto R., Uchida K., Qok C., Ruppert J., Matsuzaki J.,

RT "The Escherichia coli biotin biosynthetic enzyme sequence predicted

RT from the nucleotide sequence of the bio operon."

RT J. Biol. Chem. 263:19577-19585(1988).

RL [2]

RN SEQUENCE FROM N.A.

RP Pearson B.M., McKee R.A.;

RA "Genetic material for expression of biotin synthetase enzymes."

RT Patent number GB2216530, 11-OCT-1989.

RL [3]

RN SEQUENCE FROM N.A.

RP STRAIN-K12 / MG1655;

RC MEDLINE; 97426617.

RX Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,

RA Riley M., Collado-Vides J., Glasner F.D., Rode C.K., Mayhew G.F.,

RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,

RA Mau B., Shao Y.;

RA "The complete genome sequence of Escherichia coli K-12."

RT Science 277:1433-1474(1997).

RL -1- CATALYTIC ACTIVITY: 6-CARBOXYHEXANOYL-COA + L-ALANINE -

CC 8-AMINO-7-OXONONANOATE + COA + CO(2).

CC -1- COFACTOR: PYRIDOXAL PHOSPHATE.

CC -1- PATHWAY: BIOTIN BIOSYNTHESIS PATHWAY.

CC -1- SIMILARITY: BELONGS TO CLASS-II OF PYRIDOXAL-PHOSPHATE-DEPENDENT

CC AMINOTRANSFERASES

CC This SWISS-PROT entry is copyright. It is produced through a colla

CC between the Swiss Institute of Bioinformatics and the EMBL outc

CC the European Bioinformatics Institute. There are no restrictions

CC use by non-profit institutions as long as its content is in

CC modified and this statement is not removed. Usage and for c

CC entities requires a license agreement (See http://www.isb-sib.ch/

CC or send an email to license@isb-sib.ch).

CC --

CC EMBL: J04423; AAA23516.1;

CC EMBL: A11542; CAA00968.1;

CC EMBL: A500180; AAC73863.1;

CC PIR: D34025; SVECKP.

CC ECGENE; EG10121; BIOF.

CC PROSITE; P00599; AA-TRANSFER_CLASS_2; 1.

CC PFAM; PF00222; aminotran.2; 1.

CC Biotin biosynthesis; Transferase; pyridoxal phosphate.

KW


```

DR PIR: A53824; A53824.
DR SGD; L0001464; POMi52.
KW Nuclear protein; Transmembrane; Repeat; Glycoprotein.
FT DOMAIN 1 175 PORE SIDE (POTENTIAL).
FT TRANSMEM 176 195
FT DOMAIN 196 1337 POTENTIAL.
FT DOMAIN 196 1337 CISTERNAL SIDE (POTENTIAL).
FT CARBOHYD 280 280
FT DOMAIN 390 1276
FT REPEAT 390 413
FT REPEAT 626 650
FT REPEAT 732 755
FT REPEAT 836 859
FT REPEAT 943 966
FT REPEAT 1058 1077
FT REPEAT 1157 1178
FT REPEAT 1253 1276
SQ SEQUENCE 1337 AA; 151651 MW; A024R442069193898 CRC64;

Query Match
Best Local Similarity 7.2%; Score 81.5; DB 1; Length 1337;
Matches 45; Conservative 28; Mismatches 72; Indels 29; Gaps

QY 56 VSSLOGEDCDRGVQLGVSANLPEEQIGA---LLAGHHTLLQALRLPPTSLSKPDTRDQ 112
DB IASITGTGSDN-RCIGDSNVSFETQGVPPMKLAYSKIYNGQTFYSVDSLSQPEYFESP 437
QY 113 LQELICIPDLV-GDLASVVEGSRPL---LDSVAQQGAWLPHVADFWRVDVAITSAL 168
DB LQSSKSKQSQTQGEIANDLKWGRNQPNVINDSSITQD-----KFAKIDKI--TDGL 488
QY 169 ARSLQ-PSVLMLQLKSDGSAYRF---EVTAKFEQ-----LRYSVALLVKEM 211
DB GNVVDFTSPEELKRYDYSYNFNVHVPRAALEERFDPKSPTKRSIAIVFEI 542

RESULT 9
CH60_NEIME STANDARD; PRT: 544 AA.
AC P42385;
AT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 15-FEB-2000 (Rel. 39, Last annotation update)
DE 60 KD CHAPERONIN (PROTEIN CPN60) (GROEL PROTEIN) (63 KD STRESS
DE PROTEIN) (GSP63)
GN MOPA OR GROEL OR HSP63.
OS Neisseria meningitidis.
OC Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-2996;
RC MEDLINE; 95264914.
RA Pannekoek Y., Dankert J., van Putten J.P.M.;
RT "Construction of recombinant neisserial Hsp60 proteins and mapping of
RT antigenic domains.";
RL Mol. Microbiol. 15:277-285(1995).
CC -!- FUNCTION: PREVENTS MISFOLDING AND PROMOTES THE REFOLDING AND
CC PROPER ASSEMBLY OF UNFOLDED POLYPEPTIDES GENERATED UNDER STRESS
CC CONDITIONS (BY SIMILARITY).
CC -!- SUBUNIT: OLIGOMER OF 14 SUBUNITS COMPOSED OF TWO STACKED RINGS OF
CC 7 SUBUNITS (BY SIMILARITY).
CC -!- SUBCELLULAR LOCATION: CYTOPLASMIC AND LOOSELY ASSOCIATED WITH THE
CC OUTER MEMBRANE.
CC -!- SIMILARITY: BELONGS TO THE CHAPERONIN (HSP60) FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collabora-
CC tion between the Swiss Institute of Bioinformatics and the EMBL outstat-
CC the European Bioinformatics Institute. There are no restrictions on
CC use by non-profit institutions as long as its content is in no
CC modified and this statement is not removed. Usage by and for commer-
CC entities requires a license agreement (see http://www.isb-sib.ch/annou-
CC or send an email to license@isb-sib.ch).

```

```

DR EMBL; 222956; CAA80532.1; -.
DR HSP; P06139; 1GRL.
DR PRINTS; PR00298; CHAPERONIN60.
DR PRINTS; PR00304; TCOMPLEXCP1.
DR PRINTS; PR00296; CHAPERONINS_CPN60; 1.
DR PROSITE; PS00296; CHAPERONINS_CPN60; 1.
DR PFAM; PF00118; cpn60_TCP1; 1.
KW Chaperone; ATP-binding; Antigen.
SQ SEQUENCE 544 AA; 57580 MW; B0789DA091149D57 CRC64;

Query Match
Best Local Similarity 7.2%; Score 81; DB 1; Length 544;
Matches 61; Conservative 45; Mismatches 84; Indels 116; Gaps 14;

QY 25 LGAQLPPEVAAMAR-LLGD-----LDRSTFRKLLKFFVSSLOQEDCRDGVRL----- 72
DB 69 MGAQVYKVASNTDVGAGTTATVLAQSIIVAGMKYVTAGMNPDTLKRGDRAVAALV 128
QY 72 -----GVSNLPPEOLGALLA-----GMHTL----- 93
DB 129 EELKNIAKPCDTSKEIAVGSISAN-SDQVGAIITAEAMEKVGKGVITVEDGKLENEL 187
QY 93 -LQAALRLPPTSLKP-----DTRDQLQE----- 116
DB 199 DVVEGQFDRGYLSPIYFINDAEKQIAGLDNPFVLLFEKKISNIRDLLPVLEKVAKASRPL 247
QY 116 LCTPDVLGD-LASVVGSGORPLDLSVAQOQGANLPHVADFRRV--DVAISTALARS 172
DB 248 LIAENVEGEALATLVNNIRGILKTVAVK-----APFGDRKAMLDIAITLGTGTIVSE 303
QY 173 QPSVLMQLKLSG--GSAYRFEVP-----TAKFOELRYSVALLVKEMADL 214
DB 304 EVGLFEKATLDNLGAKRIEIKGKNTIIDGFGDAAQIEARFAELRQIETATSDY-DK 362
QY 215 EKRCR 220
DB 363 EKLQR 368

RESULT 10
PFES_PSEAE STANDARD; PRT; 446 AA.
AC Q04804;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE SENSOR PROTEIN PFES (EC 2.7.3.-).
GN PFES.
OS Pseudomonas aeruginosa.
OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonas group;
OC Pseudomonas.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=PAO;
RX MEDLINE; 93368425.
RA Dean C.R., Poole K.;
RT "Expression of the ferric enterobactin receptor (PfeA) of Pseudomonas
aeruginosa: involvement of a two-component regulatory system.";
RL Mol. Microbiol. 8:1095-1103(1993).
CC -1- FUNCTION: MEMBER OF THE TWO-COMPONENT REGULATORY SYSTEM PFES/PFES.
CC MAY ACTIVATE PFER BY PHOSPHORYLATION.
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. INNER MEMBRANE
CC (POTENTIAL).
CC -1- SIMILARITY: TO OTHER PROKARYOTIC SENSORY TRANSDUCTION HISTIDINE
CC KINASES.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL-outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).

```

```

CC EMBL; L07739; AAA25930.1; -.
DR PFAM; PF00512; signal; 1.
DR PFAM; PF00672; DUF5; 1.
KW Sensory transduction; Transferase; Kinase; Phosphorylation;
KW Transmembrane; Inner membrane.
FT DOMAIN 1 9 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 10 30 POTENTIAL.
FT DOMAIN 31 155 PERIPLASMIC (POTENTIAL).
FT TRANSMEM 156 176 POTENTIAL.
FT DOMAIN 177 446 CYTOPLASMIC (POTENTIAL).
FT MOD_RES 244 244 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
SQ SEQUENCE 446 AA; 50597 MW; 62B78FED1B4FE73E CRC64;

Query Match
Best Local Similarity 7.1%; Score 80.5; DB 1; Length 446;
Matches 46; Conservative 18; Mismatches 62; Indels 53; Gaps 8;

QY 27 AOLPP-----EVAAMARLLGD-LDRS-----TFRKLLKFFVSSLOQEDCRD 66
DB 259 SELPPEQLRORLDRICDMQRLLEDITLDLAWMDTERPQLPTEPVLESLVWEALRDDACFE 318
QY 67 G-----VORLGVSNLPPEQLGALLAGHHTLLQALRLPP-----TSLK----- 106
DB 319 SGMDPARLCRLGDCRV-EVHLDLSLAQAMENLLRNARHSHPEDGTVSLDGEREGDFWHL 377
QY 106 -----PDTFRDQLQELCIPQDLVGLDLSVFGSQRPPLDLSVAQOQGA----- 148
DB 378 RLQDQSGVGAEDQLERITFLPYRLDLSAGSGFGLGIARRAIELQGGRLWASNGKPGLC 437
QY 148 ---WLPHVA 153
DB 438 LHLWLPAAA 446

RESULT 11
BGLG_ECOLI STANDARD; PRT; 278 AA.
ID BGLG_ECOLI
AC P11989;
DT 01-OCT-1989 (Rel. 12, Created)
DT 01-OCT-1989 (Rel. 12, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE CRYPTIC BETA-GLUCOSIDE BGL OPERON ANTITERMINATOR.
GN BGLG OR BGLC.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=K12;
RX MEDLINE; 87222180.
RA Schnetz K., Toloczky C., Rak B.;
RT "Beta-glucoside (bgl) operon of Escherichia coli K-12: nucleotide
sequence, genetic organization, and possible evolutionary
relationship to regulatory components of two Bacillus subtilis
genes.";
RT J. Bacteriol. 169:2579-2590(1987).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=K12 / MG1655;
RX MEDLINE; 93315143.
RA Burland V.D., Plunkett G. III, Daniels D.L., Blattner F.R.;
RT "DNA sequence and analysis of 136 kilobases of the Escherichia coli
genome: organizational symmetry around the origin of replication.";
RT Genomics 16:551-561(1993).
RN [3]
RP SEQUENCE OF 1-197 FROM N.A.
RC STRAIN=K12;
RX MEDLINE; 87273510.
RA Mahadevan S., Wright A.;
RT "A bacterial gene involved in transcription antitermination:
regulation at a rho-independent terminator in the bgl operon of E.

```

RT Cell 50:485-494(1987).
 [4]
 RP REGULATION BY PHOSPHORYLATION.
 RX MEDLINE; 90341774.
 RA "Regulation of activity of a transcriptional anti-terminator in E.
 RT coli by phosphorylation in vivo."
 RL Science 249:540-542(1990).
 [5]
 RP REGULATION BY PHOSPHORYLATION.
 RX MEDLINE; 90311345.
 RA Schmetz K., Rak B.;
 RT "Beta-glucoside permease represses the bgl operon of Escherichia coli
 RT by phosphorylation of the antiterminal protein and also interacts
 RT with glucose-specific enzyme III, the key element in catabolite
 RT control."
 RL Proc. Natl. Acad. Sci. U.S.A. 87:5074-5078(1990).
 [6]
 RP RNA-BINDING.
 RX MEDLINE; 90381772.
 RA Houtman F., Diaz-Torre M.R., Wright A.;
 RT "Transcriptional antitermination in the bgl operon of E. coli is
 RT modulated by a specific RNA binding protein."
 RL Cell 62:1153-1163(1990).
 CC -!- FUNCTION: MEDIATES THE POSITIVE REGULATION OF THE BETA-GLUCOSIDE
 CC (BGL) OPERON BY FUNCTIONING AS A TRANSCRIPTIONAL ANTITERMINATOR.
 CC THIS IS A RNA-BINDING PROTEIN THAT RECOGNIZES A SPECIFIC SEQUENCE
 CC LOCATED JUST UPSTREAM OF TWO TERMINATION SITES WITHIN THE OPERON.
 CC -!- PTM: PHOSPHORYLATED AND INACTIVATED BY BGLF (II-BGL). THE DEGREE
 CC OF PHOSPHORYLATION IS DEPENDENT ON THE PRESENCE OR ABSENCE OF
 CC BETA-GLUCOSIDES WHICH ACT AS INDUCERS OF THE OPERON EXPRESSION.
 CC ADDITION OF INDUCER RESULT IN THE RAPID DEPHOSPHORYLATION OF BGLG.
 CC -!- SIMILARITY: BELONGS TO THE TRANSCRIPTIONAL ANTITERMINATOR BGLG
 CC FAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC ENBL; M16487; AAA23509.1;
 CC ENBL; L10328; AAA62074.1;
 CC ENBL; AE000449; AAC6746.1;
 CC ENBL; M17098; AAA23512.1;
 CC PIR; A27161; A27161.
 CC PIR; B25977; B25977.
 CC HSP; P15401; IAUU.
 CC ECOGENE; EG10116; BGLG.
 CC PROSITE; PS00654; ANTITERMINATORS_BGLG; 1.
 CC PFAM; PF00874; BglC_antitermin; 2.
 KW Transcription regulation; Activator; RNA-binding; Phosphorylation.
 SQ SEQUENCE 278 AA; 32097 MW; 5ACF1A14BF438B4F CRC64;

Query Match 7.1%; Score 80; DB 1; Length 278;
 Best Local Similarity 21.8%; Pred. No. 3.5;
 Matches 51; Conservative 42; Mismatches 91; Indels 50; Gaps 8;
 QY 17 SH--SGRVSLGALPPEVNAARLLGLDRLSTFRKLLKPVSSLOGECRDGVQRLGVS 74
 DB 54 SHELNRLSELLSHIPLEVNATCDRIISLAQERLGLQDSIYISLT-DHCOFAIKRFOON 112
 QY 75 ANLPEQLGALLAGMTLLQQAALIPDLPDTERDLQELICIPDLYGDLASVVGSG 134
 DB 113 VLLP----NPLLDIQLRYPKEFQLGEEAL---TIIDKRLGVLPKDEVGFTIAHLVSAQ 165
 QY 135 -----RPLL-----DSVAQQQAGMLPHVDFRWRVDAVST 166
 DB 166 MSGNMEDVAGVTOLMREMLQLIKFQSLNYQEEISLVQR--LVTHKLKLSWRILEHASIN 223

QY 167 ALARSLQPSVLM-----QLKLSGSAIRFEVPTAKFQELRYSLVALVLKE 210
 DB 224 DSDBSLQAVKNYPQAWQCAERIAIFIGLQYQKISPAEIMFLAINIERVKE 277
 RESULT 12
 MYSC_CAEEL STANDARD; PRT; 1947 AA.
 ID MYSC_CAEEL
 AC P12845;
 DT 01-OCT-1989 (Rel. 12, Created)
 DT 01-OCT-1989 (Rel. 12, Last sequence update)
 DT 15-JUL-1998 (Rel. 36, Last annotation update)
 DE MYOSIN HEAVY CHAIN C (MHC C).
 GN MYO-2.
 OS Caenorhabditis elegans.
 OC Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditidae;
 OC Rhabditina; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-BRISTOL N2;
 RX MEDLINE; 89178677.
 RA Dibb N.J., Maryama I.N., Krause M., Karn J.;
 RT "Sequence analysis of the complete Caenorhabditis elegans myosin
 RT heavy chain gene family."
 RL J. Mol. Biol. 205:603-613(1989).
 RN [2]
 RP SEQUENCE OF 1-21; 112-371 AND 1501-1772 FROM N.A.
 RC STRAIN-BRISTOL N2;
 RX MEDLINE; 85201409.
 RA Karn J., Dibb N.J., Miller D.M.;
 RT "Cloning nematode myosin genes."
 RL Cell Muscle Motil. 6:185-237(1985).
 CC -!- FUNCTION: MUSCLE CONTRACTION.
 CC
 CC -!- SUBUNIT: MUSCLE MYOSIN IS A HEXAMERIC PROTEIN THAT CONSISTS OF 2
 CC HEAVY CHAIN SUBUNITS (MHC), 2 ALKALI LIGHT CHAIN SUBUNITS (MLC)
 CC AND 2 REGULATORY LIGHT CHAIN SUBUNITS (MLC-2).
 CC -!- SUBCELLULAR LOCATION: THICK FILAMENTS OF THE MYOFIBRILS.
 CC -!- TISSUE SPECIFICITY: FOUND EXCLUSIVELY IN THE PHARYNGEAL MUSCLE.
 CC -!- DOMAIN: THE RODLIKE TAIL SEQUENCE IS HIGHLY REPETITIVE, SHOWING
 CC CYCLES OF A 28-RESIDUE REPEAT PATTERN COMPOSED OF 4 HEPTAPEPTIDES,
 CC CHARACTERISTIC FOR ALPHA-HELICAL COILED COILS.
 CC -!- PTM: TWO CYSTEINE RESIDUES IN THE S1 DOMAIN ARE SELECTIVELY
 CC ALKYLATED AND ARE REQUIRED FOR MYOSIN ATPASE ACTIVITY.
 CC -!- MISCELLANEOUS: EACH MYOSIN HEAVY CHAIN CAN BE SPLIT INTO 1 LIGHT
 CC MEROMYOSIN (LMM) AND 1 HEAVY MEROMYOSIN (HMM). IT CAN LATER BE
 CC SPLIT FURTHER INTO 2 GLOBULAR SUBFRAGMENTS (S1) AND 1 ROD-SHAPED
 CC SUBFRAGMENT (S2).
 CC -!- MISCELLANEOUS: THERE ARE FOUR DIFFERENT MYOSIN HEAVY CHAINS IN C.
 CC ELEGANS.
 CC -!- SIMILARITY: THE GLOBULAR HEAD AA SEQUENCE SHOWS A HIGH DEGREE OF
 CC SIMILARITY WITH THE GLOBULAR HEAD SEQUENCES OF MUSCLE & NONMUSCLE
 CC HEAVY CHAINS. BY CONTRAST THE ROD SEQUENCE IS LESS CONSERVED, BUT
 CC THE PERIODICITIES OF HYDROPHOBIC & CHARGED RESIDUES, WHICH DICTATE
 CC THE ALPHA-HELICAL COILED-COIL STRUCTURE, ARE CONSERVED.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC ENBL; X08066; CAA30855.1;
 CC ENBL; M37233; AAA28121.1;
 CC ENBL; M37235; AAA28122.1;
 CC ENBL; M37236; AAA28123.1;
 CC PIR; S05697; S05697.
 CC HSP; P08799; LMND.
 CC PFAM; PF00063; myosin_head; 1.
 CC PFAM; PF01576; Myosin_tail; 1.
 CC PRINTS; PR00193; MYOSINHEAVY.


```
RESULT 14
QAL5_NEUCR          STANDARD;          PRT;          918 AA.
AC P11637;
DT 01-OCT-1989 (Rel. 12, Created)
DT 01-OCT-1989 (Rel. 12, Last sequence update)
DE 01-FEB-1991 (Rel. 17, Last annotation update)
DE QUINATE REPRESSOR.
GN QA-1S.
OS Neurospora crassa.
OC Eukaryota; Fungi; Ascomycota; Pyrenomycetes; Sordariales;
OC Sordariaceae; Neurospora.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=74-OR23-1A;
RX MEDLINE; 89293848.
RA Geever R.F., Huilet L., Baum J.A., Tyler B.M., Patel V.B.,
RA Rutledge B.J., Case M.E., Giles N.H.;
RT "DNA sequence, organization and regulation of the qa gene cluster of
RT Neurospora crassa.";
RL J. Mol. Biol. 207:15-34 (1989).
CC -!- FUNCTION: REPRESSOR FOR ENZYMES AND PROTEINS OF QUINATE
CC METABOLISM.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X14603; CAA32753.1; -
DR PIR; S04255; S04255.
DR PIR; E31277; E31277.
DR PFAM; PF01487; DHquinase_I; 1.
DR PFAM; PF01488; Shikimate_DH; 1.
DR Quinate metabolism; Transcription regulation; Repressor; DNA-binding.
KW SEQUENCE 918 AA; 100580 MW; 67EDA399CBF098B2 CRC64;

Query Match          7.0%; Score 79.5; DB 1; Length 918;
Best Local Similarity 27.0%; Pred. No. 17;
Matches 40; Conservative 15; Mismatches 44; Indels 49; Gaps 6;

QY 17 SHSGRVFLCAQLPPEVAMARLLGLDRLSTFRKLKLVVSSLOGEDCRDGVORLGSAN 76
DQ 349 SESGTSPLGL-APHRASEISRVVGEIRDTVIPILHV----- 388
QY 77 LPPEQL--GALLAGMTLLOQALRLPPTSLKPTDFRDLQELCIPDQLVGDLSV----- 130
DQ 388 FPERALYEALLALTYLNHLRLAPDYLTVD-----LGLDGLGLQLTIVGTTK 439

QY 130 VFGSRPLLDVSAQQQAGWLPHVDFRW 157
DQ 440 VIGNKQ-----LAEVNSPRW 454

RESULT 15
CNRB_BOVIN          STANDARD;          PRT;          853 AA.
AC P23439;
DT 01-NOV-1991 (Rel. 20, Created)
DT 01-NOV-1991 (Rel. 20, Last sequence update)
DE 15-JUL-1998 (Rel. 36, Last annotation update)
DE ROD CGMP-SPECIFIC 3',5'-CYCLIC PHOSPHODIESTERASE BETA-SUBUNIT
DE (EC 3.1.4.17) (GMP-PDE BETA).
GN PDE6B OR PDEB.
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae; Bovinae; Bos.
OC -----
RN [1]
RP SEQUENCE FROM N.A.
RA Lipkin V.M., Khrantsov N.V., Vasilevskaya I.A., Atabekova N.V.,
RA Muradov K.G., Gubanov V.V., Li T., Johnston J.P., Volpp K.J.,
RA Applebury M.L.;
RT "Beta-subunit of bovine rod photoreceptor cGMP phosphodiesterase.
RT Comparison with the phosphodiesterase family.";
RL J. Biol. Chem. 265:12955-12959 (1990).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE; 90267525.
RA Lipkin V.M., Gubanov V.V., Khrantsov N.V., Vasilevskaya I.A.,
RA Atabekova N.V., Muradov K.G., Shuvaeva T.M., Surina E.A.,
RA Zagranchny V.E., Li T.;
RT "Cyclic GMP phosphodiesterase from bovine retina. Amino acid sequence
RT of beta-subunit and nucleotide sequence of corresponding cDNA.";
RL Bioorg. Khim. 16:118-120 (1990).
CC -!- FUNCTION: THIS PROTEIN PARTICIPATES IN PROCESSES OF
CC TRANSMISSION AND AMPLIFICATION OF THE VISUAL SIGNAL. NECESSARY
CC FOR THE FORMATION OF A FUNCTIONAL PHOSPHODIESTERASE HOLOENZYME.
CC -!- CATALYTIC ACTIVITY: GUANOSINE 3',5'-CYCLIC PHOSPHATE + H(2)O -
CC GUANOSINE 5'-PHOSPHATE.
CC -!- SUBUNIT: OLIGOMER COMPOSED OF TWO CATALYTIC CHAINS (ALPHA AND
CC BETA), AN INHIBITORY CHAIN (GAMMA) AND THE DELTA CHAIN.
CC -!- SUBCELLULAR LOCATION: MEMBRANE-ASSOCIATED.
CC -!- SIMILARITY: BELONGS TO THE CYCLIC NUCLEOTIDE PHOSPHODIESTERASE
CC FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; J05553; AAA30440.1; -
DR EMBL; X57146; CAA40436.1; -
DR PIR; A36617; A36617.
DR PIR; S19145; S19145.
DR PRINTS; PR00387; PDIESTERASE1.
DR PROSITE; PS00126; PDEASE_I; 1.
DR PFAM; PF00037; fer4; 1.
DR PFAM; PF00233; PDEase; 1.
DR PFAM; PF01590; GAF; 2.
KW Hydrolase; cGMP; Vision; Prenylation; Lipoprotein; Membrane.
FT CHAIN 1 850
FT PROPEP 851 853
FT LIPTD 850 850
FT CONFLICT 483 484
FT DE -> EQ (IN REF. 2)
SQ SEQUENCE 853 AA; 98330 MW; C4B3F22CFE7F2FB CRC64;

Query Match          7.0%; Score 79; DB 1; Length 853;
Best Local Similarity 20.7%; Pred. No. 17;
Matches 49; Conservative 36; Mismatches 66; Indels 86; Gaps 10;

QY 32 EVAAMARLLGD-----LDRSTFRKLKLVVSSLOGEDC-----RDGVORLGSANLP 78
DQ 55 ESAALFELVDQMENVNMRVVF-KILRLCSILHADRCSLFMYRQRNGVAEL----- 107
QY 79 EEOGLAGMTLLOQALRLPPTSLKPTDFRDLQELCIPDQLVGDLSVFGSORPL 138
DQ 107 -----ATRL--FSVQPDV---LEDCLVPPD-----SEIVFPLDGV 139
QY 139 DSVAAQ-----QQGAWLPHVDFRWVDVAISTALSRLQ-----PSVLMOLKLSDGS 186
DQ 140 GHVAQTKKVNVDVNMVMECPHFSFADELTDYVTRNLATPTIMNGKDVAVIMAVNKLDGP 199
QY 187 -----AYRFEVPTAKFQELRYSVALLVKEMADLEKR 217
```

Db 200 CFTSEDEDFLKYLNFGLNLKIIYHLSYLNHCETRRGQVLLWSANKVFEELTDIERQ 256

Search completed: May 15, 2000, 05:34:23
Job time: 12493 sec

Result No.	Score	Query		Length	DB	ID	Description
		Match	%				
1	96	8.5	195	11	Q63829	Q63829	mus musculus
2	93.5	8.3	676	5	Q27512	Q27512	caenorhabdi
3	89.5	7.9	202	5	Q9Y6G5	Q9Y6G5	mus sapien
4	88.5	7.8	413	2	Q08859	Q08859	klebsiella
5	86.5	7.6	332	10	Q048693	Q048693	arabidopsis
6	86.5	7.6	1212	5	Q9X239	Q9X239	rosophila
7	85.5	7.6	2182	11	Q88480	Q88480	rattus norv
8	83	7.3	218	2	Q88031	Q88031	streptomyce
9	83	7.3	418	4	Q9Y2F1	Q9Y2F1	homo sapien
10	82	7.2	992	11	Q9WDU50	Q9WDU50	mus musculus
11	82	7.2	1224	2	P95629	P95629	rhizobium m
12	81.5	7.2	800	11	Q62037	Q62037	mus musculus
13	81	7.2	145	1	Q28746	Q28746	archaeoglob
14	80.5	7.1	161	2	Q9WZJ7	Q9WZJ7	thermotoga
15	80	7.1	311	2	Q9ZG39	Q9ZG39	heliobacill
16	80	7.1	381	2	Q85846	Q85846	sphingomona
17	80	7.1	1968	5	Q20439	Q20439	caenorhabdi
18	79.5	7.0	831	5	Q94379	Q94379	rosophila
19	79.5	7.0	918	3	Q01393	Q01393	neurospora
20	79.5	7.0	1321	2	Q86653	Q86653	streptomyce

```
dbb      | : |      : : | | | |
175 LQDL-----VGKLDASKSLERATQ 194
```

01-NOV-1999 (TREMBlrel. 12, Last sequence update)


```

QY 108 TFRDQLQELCIPQDLVGLAS-VVFGSORPLDLSVAQOQG-----AWL 149
   : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 150 SPANAL--TAFEQEWQAIHDRQVFPVQPIVDSRSQLOGVEIILRHRGCOVLHPOTFL 207
   : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 150 PHV-ADFRWRVDVAISALSARLQ-----SVLMQKLSDGSAVRFVPTAKFOELR 201
   || || || || || || || || || || || || || || || || || || ||
Db 208 PHFRADYTWLLTAFVLOEAVQINNEYPGTFYFSVNPISLADSDSLRMVVEAR-QQLR 266
   : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 202 Y--SVALVLKEMAD 213
   || : : : : : : : : : : : : : : : : : : : : : : : : :
Db 267 QPEGVARLVLEYAE 280
   || : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 5
O48693 ID O48693 PRELIMINARY; PRT; 332 AA.
AC O48693;
DT 01-JUN-1998 (TREMBlrel. 06, Created)
DT 01-JUN-1998 (TREMBlrel. 06, Last sequence update)
DE 01-NOV-1999 (TREMBlrel. 12, Last annotation update)
DE F316.20 PROTEIN.
GN F316.20.
OS Arabidopsis thaliana (Mouse-ear cross).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons;
OC core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae;
OC Arabidopsis.
RN [1]
RC SEQUENCE FROM N.A.
RP STRAIN-CV, COLUMBIA;
RA FEDERSPIEL N.A., PALM C.J., CONWAY A.B., KURTZ D.B., CONWAY A.R.,
RA AU M., ARAUJO R., BUEHLER E., DEWAR K., FENG J., KIM C., LI Y.,
RA QUI O., OSBORNE B.I., SHINN P., SUN H., TORIUMI M., VYSOTSKAIA V.S.,
RA YU G., ECKER J., THEOLOGIS A., DAVIS R.W.;
RL Submitted (FEB-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL; AC002396; AAC00587.1; -
DR MENDEL; 27893; Arath.3357;27893.
SQ SEQUENCE 332 AA; 35831 MW; A0D178A9 CRC32;

```

```

Query Match 7.6%; Score 86.5; DB 10; Length 332;
Best Local Similarity 22.2%; Pred. No. 3;
Matches 51; Conservative 27; Mismatches 75; Indels 77; Gaps 8;

QY 28 QLPPEVAMARLIG---DLDRSTFRKLKF-----VSSSQEDCR-DG 67
   : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 112 ELSSEIDMGRRLDANTEVIQTIQEVAKLQDCTSEIKDDVKAVFDAFENLASKVCRIEG 171
   : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 68 VORLGSANLPEQLGALLAGMTLLQALRLPPTSLKPDPTFRDQLQELCIPQDLVGLA 127
   || : : : : : : : : : : : : : : : : : : : : : : : : :
Db 172 NQALPTSSLP-----ALEAAPMAPSKTSLPPAS--PDE----- 206
   || : : : : : : : : : : : : : : : : : : : : : : : : :
QY 128 SVVFGSORPLDLSVAQOQAWLPHVADFRWRVDVAISTAL-----ARSLQ 174
   || : : : : : : : : : : : : : : : : : : : : : : : : :
Db 206 -----SQSPSTPNVAKSKRGLLQHTQSMGLKDINSSSHSTFNGYFGNGASGSSS 260
   || : : : : : : : : : : : : : : : : : : : : : : : : :
QY 175 SVLMQKLSDGSAVRFVPTAKFOELRYVALVLKEMADLEKRCERLQD 224
   || : : : : : : : : : : : : : : : : : : : : : : : : :
Db 261 GVLGRINKEGGMVLEY-----VLLKELLNKRKLVNLOE 297
   || : : : : : : : : : : : : : : : : : : : : : : : : :

```

```

RESULT 6
Q9XZ29 ID Q9XZ29 PRELIMINARY; PRT; 1212 AA.
AC Q9XZ29;
DT 01-NOV-1999 (TREMBlrel. 12, Created)
DT 01-NOV-1999 (TREMBlrel. 12, Last sequence update)
DE 01-NOV-1999 (TREMBlrel. 12, Last annotation update)
DE HYPOTHETICAL 135.8 KD PROTEIN.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.

```

```

[1]
RN SEQUENCE FROM N.A.
RP ROBIN G.M., WAN K.H., HARVEY D., LEWIS S.E., BROKSTEIN P., TSANG G.,
RA AGBAYANI A., ARCAINA T.T., BAXTER E., BLAZEJ R.G., BUTENHOFF C.,
RA CHAMPE M., CHAVEZ C., CHEM M., DOYLE C.M., FARFAN D.E., FRISE E.,
RA GALLE R., GEORGE R.A., HARRIS N.L., HOSKINS R.A., EVANS-HOLM M.,
RA HOUSTON K.A., HUMMIST S.R., KIM E., LI P., MOSHREFI M., PACLEB J.M.,
RA PARK S., SEQUEIRA A., SEIHI H., SNIR E., SVIRSKAS R.R., WEINBURG T.,
RA CELNIKER S.E.;
RL Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF132186; AAD34774.1; -
KW Hypothetical protein.
SQ SEQUENCE 1212 AA; 135812 MW; 518E0F63 CRC32;

```

```

Query Match 7.6%; Score 86.5; DB 5; Length 1212;
Best Local Similarity 27.2%; Pred. No. 17;
Matches 56; Conservative 30; Mismatches 83; Indels 37; Gaps 10;

QY 30 PPEVAMARLIGDLDRSTFRKLKLFVVSSSQEDC-----RDGVORLGVGANLP 78
   || : : : : : : : : : : : : : : : : : : : : : : : : :
Db 813 PAAEAQARILASLEEL--EMNAQISDLQKVCPTDLDSRIRSAEGVQSLGESRTVS 870
   || : : : : : : : : : : : : : : : : : : : : : : : : :
QY 79 EEOGLAGLAGMTLLQALRLPPTSLKPD-TFRDQLQELCIPQDLVGLASVVGSORPL 137
   : : : : : : : : : : : : : : : : : : : : : : : : :
Db 871 KOLL-----XTLVQOR-RLQASSLINEQRTILDELRAQLLDRAQQQEDAA-----KRLRL 918
   || : : : : : : : : : : : : : : : : : : : : : : : : :
QY 138 LQSVAAQQGAWLPHVADFRWRVDVAISTS-----ALARS-----LQPSVLMQLKLSGSAYR 189
   || : : : : : : : : : : : : : : : : : : : : : : : : :
Db 919 LOSQHEEQ--MLAQRAYEEKVSVLIRTANQWAEARSPAEQQRNQIILELLSSREALQ 976
   || : : : : : : : : : : : : : : : : : : : : : : : : :
QY 190 FEVPTAKFOELRYVALVLKEMADLE 215
   || : : : : : : : : : : : : : : : : : : : : : : : : :
Db 977 QELDKLRANKSKSA-VKSEPODLD 1001
   || : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 7
O88480 ID O88480 PRELIMINARY; PRT; 2182 AA.
AC O88480;
DT 01-NOV-1998 (TREMBlrel. 08, Created)
DT 01-NOV-1998 (TREMBlrel. 08, Last sequence update)
DE 01-NOV-1998 (TREMBlrel. 08, Last annotation update)
DE CAIN.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
RN [1]
RC SEQUENCE FROM N.A.
RP MEDLINE; 98325042.
RA LAI M.M., BURNETT P.E., WOLOSKE H., BLACKSHAW S., SNYDER S.H.;
RT "Cain, a novel physiologic protein inhibitor of calcineurin.";
RL J. Biol. Chem. 273:18325-18331(1998).
DR EMBL; AF061947; AAC40176.1; -
SQ SEQUENCE 2182 AA; 242811 MW; F4CC8540 CRC32;

```

```

Query Match 7.6%; Score 85.5; DB 11; Length 2182;
Best Local Similarity 20.8%; Pred. No. 45;
Matches 49; Conservative 34; Mismatches 90; Indels 63; Gaps 8;

QY 10 YLHFGDSSHGRVSFLGAQLPPEV-----AAMARLLDGLDRSTFRK-LLKFFV 56
   || : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1236 YLHEEAARYPKKIHY--HNPELAMEALEVYFRHLHASILLKLLGKPDGVSSEVLVSPWK 1292
   || : : : : : : : : : : : : : : : : : : : : : : : : :
QY 57 SSLQ-----EDCRDGVORL-GVSANLPEQLGALLAGMTL 92
   : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1293 EAAEGPFARGEKNTPKASEKEKACLVDESDSHSSAGTLPQGCASLPSSSGPGLTSPPYTA 1352
   || : : : : : : : : : : : : : : : : : : : : : : : : :
QY 93 L-----QOALRLPPTSLKPDPTFRDQLQELCIPQDLVGLASVVGSORPL-- 138
   || : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1353 TPIHDHYVKCKKPRQQA---TPDRSQBSTAVALSDSSSTQDFFNEPTSLDGSRLKLP 1409
   || : : : : : : : : : : : : : : : : : : : : : : : : :

```

```

QY 138 -----LDSVAOQOAWLPHVADFRWRVDAISTALARSLOPSVLMQLKLSGDSAY 188
DB 1410 KRISLSAQAGSPGKDLPGPTTEERGKTEESLESTAEFRVFEVQKPKVADSSASAY 1465

RESULT 8
O88031 PRELIMINARY; PRT; 218 AA.
AC O88031;
DT 01-NOV-1998 (TremBLrel. 08, Created)
DT 01-NOV-1998 (TremBLrel. 08, Last sequence update)
DT 01-NOV-1998 (TremBLrel. 08, Last annotation update)
DE HYPOTHETICAL 23.7 KD PROTEIN.
GN SC5A7_25C.
OS Streptomyces coelicolor.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
[1]
RP SEQUENCE FROM N.A.
RC STRAIN-A3(2);
RA SEGER K.J., HARRIS D.; to the EMBL/GenBank/DBJ databases.
[2]
RP SEQUENCE FROM N.A.
RC STRAIN-A3(2);
RA PARKHILL J., BARRELL B.G., RAJANDREAM M.A.;
RL Submitted (JUL-1998) to the EMBL/GenBank/DBJ databases.
[3]
RP SEQUENCE FROM N.A.
RC STRAIN-A3(2);
RX MEDLINE; 97000351.
RA KINASHI H., HOPWOOD D.A.;
RA REDENBACH M., KIESER H.M., DENAPAITTE D., EICHNER A., CULLUM J.,
RT "A set of ordered cosmids and a detailed genetic and physical map for
RT the 8 Mb Streptomyces coelicolor A3(2) chromosome."
RT Mol. Microbiol. 21:77-96(1996).
DR EMBL; AL031107; CAA19954.1; -.
KW Hypothetical protein.
SQ SEQUENCE 218 AA; 23670 MW; 19769615 CRC32;

Query Match 7.3%; Score 83; DB 2; Length 218;
Best Local Similarity 27.1%; Pred. No. 3.6;
Matches 48; Conservative 22; Mismatches 51; Indels 56; Gaps 9;

QY 68 VORLGSANLPEEQALLAGMHT-----LLOQALRLPPTSLKPDTERDQLQELCIP 119
DB 40 VORLGSANLPEEQALLAGMHT-----LLOQALRLPPTSLKPDTERDQLQELCIP 98

QY 120 QDLVG-----DLASVVGSGORPLLDVAQOQAWLPHVADFRWRVDAISTALARSLOPS 175
DB 98 -DGAGMSRQOAGALINGDRRPTMEHCDAIQRFVRH-AGFL---TAEDPEALAGALQ-- 150

QY 176 VLMQLKLSGDSAYREVTAKQELRYVALVKEMADLEKRC-----RRRLQD 224
DB 150 -----RTE-----QEL-----LQALDRERQAAAPADPLERLLQD 180

RESULT 9
Q9Y2F1 PRELIMINARY; PRT; 418 AA.
AC Q9Y2F1;
DT 01-NOV-1999 (TremBLrel. 12, Created)
DT 01-NOV-1999 (TremBLrel. 12, Last sequence update)
DT 01-NOV-1999 (TremBLrel. 12, Last annotation update)
DE KIAA0942 PROTEIN (FRAGMENT).
GN KIAA0942.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Primates; Catarrhini; Hominiidae; Homo.
[1]
RP SEQUENCE FROM N.A.
RC TISSUE-BRAIN;

```

```

RX MEDLINE; 99246063.
RA NAGASE T., ISHIKAWA K., SUYAMA M., KIKUNO R., HIROSAWA M.,
RA MIYAJIMA N., TANAKA A., KOTANI H., NOMURA N., OHARA O.;
RT "Prediction of the coding sequences of unidentified human genes. XIII.
RT The complete sequences of 100 new cDNA clones from brain which code
RT for large proteins in vitro."
RL DNA Res. 6:63-70(1999).
DR EMBL; AB023159; BAA76786.1; -.
FT NON_TER 1
SQ SEQUENCE 418 AA; 47378 MW; 02662B38 CRC32;

Query Match 7.3%; Score 83; DB 4; Length 418;
Best Local Similarity 23.8%; Pred. No. 8.6;
Matches 57; Conservative 41; Mismatches 98; Indels 44; Gaps 13;

QY 1 MSVGAAT-PYLHHPGDHSG--RVSEFLGALQLPPEVAAARLLGDLDRSTFRLLKEVVS 57
DB 135 ISRIGSTTNFPLDIPHPDPAVNAVYKSGFLARKIHADMDGKTPRGKWKTFYAVLKGTVL 194
QY 58 SLOGEDCRDGVORLGSANLPEEQALLAGMHTLLOQAL---RLPPT-SLKPDTERDQL 113
DB 195 YLQDEYKP-----EKALSEEDLKNVSVHVALASKATDYKKPNVFKLTADWRVLL 247
QY 114 QELCIPQDLVG-----DLASVVF-----GSQ-----RPLDLSVA---OQOQAWLPHV 152
DB 248 FQTSPEEMQGWINKINCAVAFSAPPFPAAGSOKKFSRPLLPATTTKLSQEQLKSHE 307
QY 153 ADFRWRVDVAIST-ALARSLOPSVLMQLKLSGDSAYREVTAKQELRYVAL-VLKE 210
DB 308 SKLK-----QITTELAEHRSPPD--KKVAKDVDEYKLDHYLEFEKTRYEMVSVILKE 360

RESULT 10
Q9WU50 PRELIMINARY; PRT; 992 AA.
AC Q9WU50;
DT 01-NOV-1999 (TremBLrel. 12, Created)
DT 01-NOV-1999 (TremBLrel. 12, Last sequence update)
DT 01-NOV-1999 (TremBLrel. 12, Last annotation update)
DE DYSTONIN (FRAGMENT).
GN BPAGL-N.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
[1]
RP SEQUENCE FROM N.A.
RC STRAIN-BALB/C;
RX MEDLINE; 99268151.
RA LEUNG C.L., SUN D., LIEM R.K.H.;
RT "The intermediate filament protein peripherin is the specific
RT interaction partner of mouse BPAGL-n (dystonin) in neurons."
RL J. Cell Biol. 144:435-446(1999).
DR EMBL; AF115383; AAD22959.1; -.
FT NON_TER 1
SQ SEQUENCE 992 AA; 112950 MW; A0959F08 CRC32;

Query Match 7.2%; Score 82; DB 11; Length 992;
Best Local Similarity 22.4%; Pred. No. 33;
Matches 59; Conservative 30; Mismatches 106; Indels 68; Gaps 11;

QY 2 SAVGAATPYLHHPGDHSGHRSVFL-----GAQLPPEVAA 35
DB 223 SAFDTAGRECHHPAEISPCNSGHLNLRPLSRWTOBPHOTEGKWPRAAEQLPKVE-- 281
QY 36 MARLLG-DLDRSTFRLLKLVVSVSLOGE-----DCRQGVORLGSANLPEEQALLAGM 89
DB 281 QFRQGPALDRESSQPCYSEVFSQTSTELQITFDKKNPITR-----LSEL 325
QY 90 HTLLOQALR--LPPTSLKPDTERDQLQELCIPQDLV-----GDLASVVGSGORPLDLSVA 142
DB 326 ETMREQALHPSRPVTVYQDDKLERELVLLTPLEIAKNKQCGMHTVTTLAKQERLGSAA 385

```

QY 143 QOQAWLPHVADFRVDAVIASTALSARLSQPSVLMQLKSDGSAYRFPVPTAKFOELRY 202
Db 386 ---GGHMLGEC---RTSGGLKGDFLKKSVEPEASPSLDLNOACSVRDE--EFQFQGLRH 436
QY 203 SVA---LVLMKEMADLEKRCERRL 222
Db 437 TVTGRQVLEAKLLDMRTVBOURL 459

RESULT 11
P95629 PRELIMINARY; PRT; 1224 AA.
AC P95629;
DT 01-MAY-1997 (TREMBlrel. 03, Created)
DT 01-MAY-1997 (TREMBlrel. 03, Last sequence update)
DT 01-NOV-1999 (TREMBlrel. 12, Last annotation update)
DE PUTA GENE.
GN PUTA.
OS Rhizobium meliloti.
OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
OC Rhizobiaceae; Sinorhizobium.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-GR4;
RX MEDLINE; 97157944.
RA JIMENEZ-ZURDO J.I., GARCIA-RODRIGUEZ F.M., TORO N.;
RT "The Rhizobium meliloti puta gene: its role in the establishment of
RL the symbiotic interaction with alfalfa";
RL Mol. Microbiol. 23:85-93(1997).
DR EMBL; Y08500; CAA69727.1; -;
DR PFAM; YF00171; aldedh; 1.
DR PFAM; PF01619; Pro-dh; 1.
SQ SEQUENCE 1224 AA; 131190 MW; B2C95AE0 CRC32;

Query Match 7.28; Score 82; DB 2; Length 1224;
Best Local Similarity 28.1%; Pred. No. 44;
Matches 25; Conservative 21; Mismatches 33; Indels 10; Gaps 3;
QY 29 LPP--EVAAMARLLGLDRLSTFRKLLKFWVSSLOGEDCRDGVORLGVSANLPFOLGALL 86
Db 47 LPPLVEAATOSKEIRDAASTARKL---TEALRGKSGSGVEGLVOEYSLSSQEGVALM 102
QY 87 AGMHTLQQALRLPPTSLKPDTPRDLQOE 115
Db 103 C----LAERPVRIPDTATRALIRDKIAD 127

RESULT 12
Q62037 PRELIMINARY; PRT; 800 AA.
AC Q62037;
DT 01-NOV-1996 (TREMBlrel. 01, Created)
DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
DT 01-NOV-1999 (TREMBlrel. 12, Last annotation update)
DE ROD PHOSPHODIESTERASE BETA SUBUNIT ISOZYME.
OS Mus musculus (Mouse)
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=RETINA;
RX MEDLINE; 91130581.
RA BAEHR W., CHAMPAGNE M.S., LEE A.K., PITTIER S.J.;
RT "Complete cDNA sequences of mouse rod photoreceptor cGMP
RT phosphodiesterase alpha- and beta-subunits, and identification of
RT beta", a putative beta-subunit isozyme produced by alternative
RT splicing of the beta-subunit gene";
RL FEBS Lett. 278:107-114(1991).
DR EMBL; X87952; CAA61202.1; -;
DR PROSITE; PS00126; PDEASE_I; 1.
DR PFAM; PF01590; GAF; 2.

DR PFAM; PF00233; PDease; 1.
DR PRINTS; PRO0387; PDIESTERASE1.
SQ SEQUENCE 800 AA; 92076 MW; 3204E9D4 CRC32;
Query Match 7.28; Score 81.5; DB 11; Length 800;
Best Local Similarity 19.4%; Pred. No. 28;
Matches 49; Conservative 38; Mismatches 68; Indels 97; Gaps 10;
QY 32 EVAAMARLLGLDRLS-----TFRKLLKFWVSSLOGEDC-----RDGVORLGVSANLPE 79
Db 55 ESAAFLVQDMQESVNMNERVFKILLRLCTILHADRCSLFMYRQRNGIAEL----- 107
QY 80 EQLGALLAGMHTLQQALRLPPTSLKPDTPRDLQOEICIPQDLVGLDASVVFSGRPLLD 139
Db 107 -----ATRL--FSVQPSDL--LEDCLVPPD-----SEVFFLDIGIVG 140
QY 140 SVAQ-----QOQAWLPHVADFRVDAVIASTALSARLSQ-----PSVLMQLKSDGS- 187
Db 141 HVAQTKMINVQDVACPFHSSFADELTDYVTKNILSTFIMNGKDVAVIMAVNKLDPGC 200
QY 187 -----AYRFEVPTAKFOELRYSLVALVKEMADLEKR----- 218
Db 201 FTSEDEDVFTKYLNFATLNKLIYHLSYLHNCETRRGOVLWSANKVFEUTDIEROFHKA 260
QY 218 -----CER 220
Db 261 FYTVRAYLNCER 272

RESULT 13
O28746 PRELIMINARY; PRT; 145 AA.
ID O28746;
AC O28746;
DT 01-JAN-1998 (TREMBlrel. 05, Created)
DT 01-JAN-1998 (TREMBlrel. 05, Last sequence update)
DT 01-NOV-1999 (TREMBlrel. 12, Last annotation update)
DE CONSERVED HYPOTHETICAL PROTEIN.
GN AF1526.
OS Archaeoglobus fulgidus.
OC Archaea; Euryarchaeota; Archaeoglobales; Archaeoglobaceae;
OC Archaeoglobus.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=VC-16 / DSM 4304 / ATCC 49558;
RX MEDLINE; 98049343.
RA KLENK H.-P., CLAYTON R.A., TOMB J.-F., WHITE O., NELSON K.E.,
RA KETCHUM K.A., DODSON R.J., GWINN M., HICKEY E.K., PETERSON J.D.,
RA RICHARDSON D.L., KERLAVAGE A.R., GRAHAM D.E., KYRIDES N.C.,
RA FLEISCHMANN R.D., QUACKENBUSH J., LEE N.H., SUTTON G.G., GILL S.,
RA KIRKNESS E.F., DOUGHERTY B.A., MCKENNEY K., ADAMS M.D., LOFTUS S.,
RA PETERSON S., REICH C.I., MCNEIL L.K., BADGER J.H., GLODEK A., ZHOU L.,
RA OVERBEK R., GOCAYNE J.D., WEIDMAN J.F., MCDONALD L., UTTERBACK T.,
RA COTTON M.D., SPRIGGS T., AFTIACH P., KATNE B.P., SYKES S.M.,
RA SADOW P.W., D'ANDREA K.P., BOWMAN C., FUJII C., GARLAND S.A.,
RA MASON T.M., OLSEN G.J., FRASER C.M., SMITH H.O., WOESE C.R.,
RA VENTER J.C.;
RT "The complete genome sequence of the hyperthermophilic, sulphate-
RT reducing archaeon Archaeoglobus fulgidus";
RL Nature 390:364-370(1997).
DR EMBL; AE000997; AAB89722.1; -;
DR HSSP; Q57997; 1M3H.
DR TIGR; AF1526; -;
KW Hypothetical protein.
SQ SEQUENCE 145 AA; 16339 MW; E5041F32 CRC32;

Query Match 7.28; Score 81; DB 1; Length 145;
Best Local Similarity 30.7%; Pred. No. 3.2;
Matches 35; Conservative 21; Mismatches 40; Indels 18; Gaps 5;
QY 36 MARLLGLDRLS-----TFRKLLKFWVSSLOGEDCRDGVORLGVSANLPFOLGALLGMH 90
Db 103 C----LAERPVRIPDTATRALIRDKIAD 127

RT "Tracking molecular evolution of photosynthesis by characterization of
 RT a major photosynthesis gene cluster from *Helicobacillus mobilis*,"
 RL Proc. Natl. Acad. Sci. U.S.A. 95:14851-14856(1998).
 DR EMBL; AF080002; AAC84010.1;
 FT NON_TER 1
 SQ SEQUENCE 311 AA; 33883 MW; B91E34DE CRC32;

Query Match 7.1% Score 80; DB 2; Length 311;
 Best Local Similarity 22.0%; Pred. No. 11;
 Matches 59; Conservative 32; Mismatches 81; Indels 96; Gaps 14;
 QY 5 GAATPVLLHHPGDSHSGRVSLGAGLPPEVAAMARLLGLDLSR-----TFKLLKEV 55
 DB 27 GKQTPVLLHHPGDSHSGRVSLGAGLPPEVAAMARLLGLDLSR-----TFKLLKEV 55
 QY 56 VS-----SLOGED-CRDGVQRLGVSANLPEEQGALLAGM---HTLQOALRLPPTSLKPD 107
 DB 85 TSAVREALNGQDFCRRVYERFLEV-----ALLSGTEGHLSTYQGAVERPEGLAAG 136
 QY 108 TFRDQLQELCIPDQDLASVVGSGORPLLDSSVAQQGAWLPHVADFRW----- 158
 DB 137 -----RIPVVVDIGGSAEVF-----HHKDRWRQSFPLGAVR 170
 QY 158 -----RVDVAI-----STSLARSLOP-----SVLMQLK-LSDGSAYR 189
 DB 171 LTESPLGRVDIASVWAPAVEKVSLSQRMGRPILIGVGGTITTTVAATLQRLRDYAPDKVHG 230
 QY 190 FEVPTAKFQELRYSLVALVLEKEMADLEKR 217
 DB 231 YEIPLEKVK-----SIARELRAM-DLEER 253

Search completed: May 15, 2000, 05:33:29
 Job time: 12455 sec

DB 1 MERILLVDDTGRGEIAPQKLLKLAEDGLRGVEVILYIREMEVPPFVPEEK---ELAAHY 57
 QY 91 TLLOQALRLPPTSLKPDFTFRQLOELCIPQDLVGLASVVGSGORPLLDSSVAQ 144
 DB 58 RLMTOSMK-----KLEGFKNQLEKAGLK---VSDV-SVVEFGKYADRLLLVEKQ 101

RESULT 14
 Q9WZJ7 PRELIMINARY; PRT; 161 AA.
 AC Q9WZJ7;
 DT 01-NOV-1999 (Tremblrel. 12, Created)
 DT 01-NOV-1999 (Tremblrel. 12, Last sequence update)
 DT 01-NOV-1999 (Tremblrel. 12, Last annotation update)
 DE HYPOTHEICAL 18.1 KD PROTEIN.
 GN TM0738.
 OS Thermotoga maritima.
 OC Bacteria; Thermotogales; Thermotoga.
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE; 99287316.
 RA NELSON K.E., CLAYTON R.A., GILL S.R., GWINN M.L., DODSON R.J.,
 RA HAFT D.H., HICKEY E.K., PETERSON J.D., NELSON W.C., KETCHUM K.A.,
 RA MCDONALD L., UTTERBACK T.R., MALEK J.A., LINHER K.D., GARRETT M.M.,
 RA STEWART A.M., COTTON M.D., PRATT M.S., PHILLIPS C.A., RICHARDSON D.,
 RA HEIDELBERG J., SUTTON G.G., FLEISCHMANN R.D., WHITE O., SALZBERG S.L.,
 RA SMITH H.O., VENTER J.C., FRASER C.M.;
 RA Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.
 RL EMBL; AE001744; AAD35819.1;
 KW Hypothetical protein.
 SQ SEQUENCE 161 AA; 18092 MW; B3E9F115 CRC32;

Query Match 7.1%; Score 80.5; DB 2; Length 161;
 Best Local Similarity 27.7%; Pred. No. 4.1;
 Matches 26; Conservative 20; Mismatches 35; Indels 13; Gaps 3;

QY 1 MSVAGATP-YLHHPGDSHSGRVSLGAGLPPEVAAMARLLGLDLSRTRKLLKFVSSL 59
 DB 69 LAAVAGRTQVFIVTPGTSKSRTHSLGSELQDKQKQIERKILELRQ-----LKTETDPL 123
 QY 60 QGEDCRGVQRLGVSANL-----PEEQGALL 86
 DB 124 ERERLEDLERLKLALNKLKASRAPELLVGLL 157

RESULT 15
 Q9ZGG9 PRELIMINARY; PRT; 311 AA.
 AC Q9ZGG9;
 DT 01-MAY-1999 (Tremblrel. 10, Created)
 DT 01-MAY-1999 (Tremblrel. 10, Last sequence update)
 DT 01-MAY-1999 (Tremblrel. 10, Last annotation update)
 DE EXOPOLYPHOSPHATASE PPX (FRAGMENT).
 GN PPX.
 OS *Helicobacillus mobilis*.
 OC Bacteria; Firmicutes; Bacillus/Clostridium group;
 OC *Helicobacterium* group; *Helicobacillus*.
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE; 99061957.
 RA XIONG J., INOUE K., BAUER C.E.;

Mature Pseudomonas
eryA region polype
Mature Pseudomonas
Polymerase enhanci
Notch IN3K full le
Human BAZ1-beta pr
Huntingtin interac
Huntingtin interac
Angiotensin. New c
Mycobacterium tube
Mycobacterium tube

35 77.5 7.0 289 1 R88024
36 77 6.9 3567 1 R44431
37 76.5 6.9 289 1 R88022
38 76.5 6.9 403 1 W72844
39 76.5 6.9 1078 1 R28963
40 76.5 6.9 1527 1 W81172
41 76 6.8 388 1 W18029
42 76 6.8 914 1 W18030
43 75.5 6.8 477 1 P50636
44 75 6.7 344 1 W32425
45 75 6.7 344 1 W32357

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 14, 2000, 23:18:10 : Search time 38.86 Seconds
(without alignments)
136.533 Million cell updates/sec

Title: us-09-223-796-2

Perfect score: 1112
Sequence: 1 MSALGAAAPYLHPADSHG.....ALVLKMAELEKCKERKLD 224

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 188963 segs, 23696106 residues 188963

Total number of hits satisfying chosen parameters:

Minimum DB seq length: 0

Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database : A_Geneseq_36.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1112	100.0	224	1 W37723	Rat Hypertension r
2	99.5	8.9	359	1 W29380	S. pneumoniae pept
3	99.5	8.9	359	1 W38592	S. pneumoniae pept
4	86.5	7.8	195	1 W71684	Amino acid sequenc
5	86.5	7.8	196	1 W07867	Human secreted pro
6	84.5	7.6	1786	1 W24790	P. falciparum live
7	82.5	7.4	289	1 W33924	Pseudomonas pseudo
8	82.5	7.4	758	1 W46270	Moraxella catarrha
9	82.5	7.4	1196	1 W5189	Osteoinductive ret
10	82	7.4	427	1 P90476	Polypeptide of hum
11	82	7.4	536	1 R76502	Human MIS mature p
12	82	7.4	559	1 P70196	Sequence encoded b
13	82	7.4	560	1 R76501	Human MIS protein.
14	81	7.3	3588	1 R34712	Bacillus subtilis
15	80.5	7.2	289	1 R88023	Mature Pseudomonas
16	80	7.2	1048	1 W97720	Staphylococcus aur
17	79.5	7.1	289	1 R88021	Mature Pseudomonas
18	79.5	7.1	875	1 W48309	Pisum sativum ACCa
19	79	7.1	660	1 Y13349	Amino acid sequenc
20	79	7.1	1145	1 W32097	Miniature swine re
21	79	7.1	1194	1 W32722	Porcine retrovirus
22	78.5	7.1	169	1 W88390	Human Zneul partia
23	78.5	7.1	181	1 W88391	Human Zneul partia
24	78.5	7.1	254	1 W88382	Human neuro-growth
25	78.5	7.1	273	1 W88381	Human neuro-growth
26	78.5	7.1	363	1 W06320	Human mitogen-acti
27	78.5	7.1	363	1 W97670	Human mitogen-acti
28	78.5	7.1	393	1 W06321	Human mitogen-acti
29	78.5	7.1	393	1 W97671	Human mitogen-acti
30	78.5	7.1	399	1 W06322	Human mitogen-acti
31	78.5	7.1	399	1 W97672	Human mitogen-acti
32	78.5	7.1	1427	1 R10534	Human 160kD mediat
33	78	7.0	687	1 W41586	Truncated restin p
34	77.5	7.0	289	1 R88025	Mature Pseudomonas

ALIGNMENTS

RESULT 1
W37723
ID W37723 standard; Protein; 224 AA.
AC W37723:
DT 09-JUN-1998 (first entry)
DE Rat Hypertension related calcium regulated.
KW Hypertension related calcium regulation; antibody; hypertension;
KW extracellular calcium concentration; asthma; diabetes; stroke;
KW hyperthyroidism; osteoporosis; heart failure; diabetes; stroke;
KW cancer; inflammatory disease; asthma.
OS Rattus rattus. Location/Qualifiers
FH Key
FT Misc_difference 15..21 /note= "EF-hand like motif"
PN WO9749807-A2.
PD 31-DEC-1997.
PF 23-JUN-1997; CA0439.
PR 21-JUN-1996; US-667495.
PA (GOS)/ GOSARD F.
PA (HAME)/ HAMET P.
PA (LEWA)/ LEWANCZUK R.
PA (TREM)/ TREMBLAY J.
PI Gossard F, Hamet P, Lewanczuk R, Tremblay J.
DI WPI: 98-077171/07.
DR N-PSDB: V18890.
PT Hypertension related calcium regulated gene - useful to develop
PT products to treat or detect, e.g. hypertension, stroke,
PT osteoporosis, heart failure, cancer, diabetes or asthma
PS Claim 8; Pages 26-27; 46pp; English.
CC This is the amino acid sequence of the hypertension related calcium
CC regulated gene (HARG), which was isolated from the rat parathyroid.
CC Its expression is regulated by extracellular calcium concentration.
CC An antibody against the protein, can be used to detect or modulate
CC (e.g. enhance or inhibit) abnormal calcium levels. They can
CC specifically be used to detect or treat, e.g. hypertension,
CC hyperthyroidism, osteoporosis, heart failure, diabetes, cancer,
CC inflammatory disease, and asthma.
SQ Sequence 224 AA;

Query Match 100.0%; Score 1112; DB 1; Length 224;
Best Local Similarity 100.0%; Pred. No. 1.1e-101;
Matches 224; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MSALGAAAPYLHPADSHSGRVFLGSGPSPEVTAVQAQLKDLDRSTFRKLLKLVGALH 60
Db 1 MSALGAAAPYLHPADSHSGRVFLGSGPSPEVTAVQAQLKDLDRSTFRKLLKLVGALH 60
QY 61 GKDCREAVQELGASANTSEERLAVLLAGTHTLLOQALRLPPASLKPDFAFQEEQLGIPQ 120
Db 61 GKDCREAVQELGASANTSEERLAVLLAGTHTLLOQALRLPPASLKPDFAFQEEQLGIPQ 120
QY 121 DLIGDLSLAGSGSORPLDSDVAQGGSSLPVSVFRNRVDVAISTSAQSRSLOPVSVMQL 180
Db 121 DLIGDLSLAGSGSORPLDSDVAQGGSSLPVSVFRNRVDVAISTSAQSRSLOPVSVMQL 180
QY 181 KLTGSAHRFEVPIAKFOELRYSVALVKMAELEKCKERKLD 224

```

Db 181 KLTDGSAHFEVPIAKFOELRYSAVLVKEMAELEKCKERKLD 224
|||||
RESULT 2
W29380
ID W29380 standard; Protein; 359 AA.
AC W29380;
DT 20-MAY-1998 (first entry)
DE S. pneumoniae peptide releasing factor RF-1.
KW RF-1; peptide releasing factor; Streptococcus pneumoniae; diagnosis;
KW otitis media; conjunctivitis; meningitis; pneumonia; endocarditis;
KW bacterial infection.
OS Streptococcus pneumoniae.
PN W09743304-A1.
PD 20-NOV-1997.
PF 14-MAY-1997; U08272.
PR 14-MAY-1996; US-017670.
PA (SMIK ) SMITHKLINE BEECHAM CORP.
PA (SMIK ) SMITHKLINE BEECHAM PLC.
PI Black MT, Hodgson JE, Knowles DJC, Nicholas RO,
PI Stodola RK;
DR WPI: 98-008794/01.
DR N-PSDB; T89027.
DT DNA encoding peptide releasing factor, RF-1 from Streptococcus
PT pneumoniae - useful for diagnosis and treatment of, e.g. otitis
PT media, conjunctivitis and meningitis, etc.
PS Claim 12; Page 9; 41pp; English.
CC This is a peptide releasing factor (prfa) RF-1 from Streptococcus
CC pneumoniae. RF-1 is one of the three factors required for the
CC termination of protein synthesis in bacteria. Upon recognition of stop
CC codons, RF-1 binds to the ribosome resulting in the peptidyl transferase
CC centre of the ribosome switching to a hydrolytic function to remove the
CC completed polypeptide chain. A mutation in S. typhimurium prfa has been
CC demonstrated to inhibit cell division through a novel regulatory circuit.
CC Hence, RF-1 may be used as an anti-bacterial target. RF-1 can be used to
CC identify compounds which interact with and inhibit or activate its
CC activity. RF-1 antagonists can be used to treat diseases caused by
CC S. pneumoniae. Through genetic immunisation, immunological response in a
CC mammal can be induced by inoculation with RF-1 or delivery of RF-1
CC encoding nucleic acid in a vector, adequate to produce antibody and/or
CC T cell immune responses to protect the animal from disease. The products
CC and methods are particularly useful for diagnosis of disease, preferably,
CC bacterial infections caused by S. pneumoniae, especially otitis media,
CC conjunctivitis, pneumonia, bacteremia, meningitis, sinusitis, pleural
CC empyema and endocarditis.
SQ Sequence 359 AA;

Query Match 8.9%; Score 99.5; DB 1; Length 359;
Best Local Similarity 24.6%; Pred. No. 0.054;
Matches 49; Conservative 37; Mismatches 58; Indels 55; Gaps 11;

QY 48 FRKLLVVGALHGKDCRAVEOLGASANLSEERLAVLAGHTLLQQALRPPASLRPD 107
Db 53 YKQVLQNV-----DAEEMIKESGGADLEE-----LAKQELK--DAKAEKE 92
QY 108 AFQEEQLQELGIPQD-----LI-----GDLASLAFSGSRPLDLSVAQOQG----- 147
Db 93 EYEELKILLPKDPNDKNIIIEIRGAGGDEAALFAGDLTMTQKYAEAGQWRFEYME 152
QY 147 SSLPHVSYFRWRVDAVISTASQSRSLQPSVLQMLKLTGCSAHRFE-VPIAKFQ---ELRY 202
Db 153 ASNMGVGGFK---EVVAMVSGQ-----SVYSKLKYESG-AHRQVRVPVTESQGRVHTST 202
QY 203 SVALVKEMAELEKCKERC 221
Db 203 ATVLVMPVEVEYDIDPK 221

RESULT 3
W38592
ID W38592 standard; Protein; 359 AA.

```

```

AC W38592;
DT 09-NOV-1998 (first entry)
DE S. pneumoniae peptide chain release factor 1.
KW Streptococcus pneumoniae protein; genetic immunisation; antagonist;
KW immunological response; inoculation; antibody production; inhibitor;
KW T cell immune response; antimicrobial compound; bacterial adhesion;
KW extracellular matrix protein; protein-mediated cell invasion; wound;
KW pathogenesis.
OS Streptococcus pneumoniae.
PN W09743303-A1.
PD 20-NOV-1997.
PF 14-MAY-1997; U07950.
PR 14-MAY-1996; US-017670.
PA (SMIK ) SMITHKLINE BEECHAM CORP.
PA (SMIK ) SMITHKLINE BEECHAM PLC.
PI Black MT, Hodgson JE, Knowles DJC, Nicholas RO,
PI Stodola RK;
DR WPI: 98-008793/01.
DR N-PSDB; T98642.
DT Novel Streptococcus pneumoniae proteins and related DNA - useful for
PT diagnosing anti-microbial agents for treatment of bacterial
PT infections
PS Claim 12; Pages 359-360; 483pp; English.
CC This sequence represents a Streptococcus pneumoniae protein that, based
CC on homology with a Bacillus subtilis protein, is a peptide chain release
CC factor 1, and is encoded by a DNA sequence of the invention.
CC The DNA sequences were isolated from Streptococcus pneumoniae strain
CC 0100993 (NCIMB 40794). The Streptococcus pneumoniae proteins of the
CC invention can be used to identify compounds which interact with and
CC inhibit or activate the activity of the proteins. Antagonists can be
CC used to treat diseases caused by S. pneumoniae proteins, through genetic
CC immunisation. They can also be used to induce an immunological response
CC in a mammal by inoculation with the S. pneumoniae proteins or delivery
CC of the encoding nucleic acids in a vector adequate to produce antibody
CC and/or T cell immune responses to protect the animal from disease. The
CC proteins can also be used to identify antimicrobial compounds which are
CC capable of inhibiting their bioactivity. In particular the proteins of
CC the invention can be used to prevent adhesion of bacteria to mammalian
CC extracellular matrix proteins on in-dwelling devices or in wounds, to
CC block protein-mediated mammalian cell invasion, and to block the normal
CC progression of pathogenesis in infections initiated other than by the
CC implantation of in-dwelling devices or other surgical techniques.
SQ Sequence 359 AA;

Query Match 8.9%; Score 99.5; DB 1; Length 359;
Best Local Similarity 24.6%; Pred. No. 0.054;
Matches 49; Conservative 37; Mismatches 58; Indels 55; Gaps 11;

QY 48 FRKLLVVGALHGKDCRAVEOLGASANLSEERLAVLAGHTLLQQALRPPASLRPD 107
Db 53 YKQVLQNV-----DAEEMIKESGGADLEE-----LAKQELK--DAKAEKE 92
QY 108 AFQEEQLQELGIPQD-----LI-----GDLASLAFSGSRPLDLSVAQOQG----- 147
Db 93 EYEELKILLPKDPNDKNIIIEIRGAGGDEAALFAGDLTMTQKYAEAGQWRFEYME 152
QY 147 SSLPHVSYFRWRVDAVISTASQSRSLQPSVLQMLKLTGCSAHRFE-VPIAKFQ---ELRY 202
Db 153 ASNMGVGGFK---EVVAMVSGQ-----SVYSKLKYESG-AHRQVRVPVTESQGRVHTST 202
QY 203 SVALVKEMAELEKCKERC 221
Db 203 ATVLVMPVEVEYDIDPK 221

RESULT 4
W71684
ID W71684 standard; Protein; 195 AA.
AC W71684;
DT 04-DEC-1998 (first entry)
DE Amino acid sequence of the human tumorigenesis associated protein.
KW Human; tumorigenesis associated protein; HTAP; transplantation;

```


KW vaccine; immunotherapy; malaria.
OS Plasmodium falciparum.
FH Key Location/Qualifiers
FT region 223..278
FT /note="repeat region 1"
FT region 279..318
FT /note="repeat region 2"
FT region 1537..1576
FT /note="repeat region 3"
PN W09641877-A2.
PD 27-DEC-1996.
PF 12-JUN-1996; F00894.
PR 13-JUN-1995; FR-007007.
PA (INSP) INST PASTEUR.
PI Daubersies P, Druilhe P;
DR WPI: 97-065464/06.
DR N-PSDB; T78868.
PT Plasmodium falciparum poly:peptide(s) and related nucleic acids -
PT derived from the liver stage antigen-3, useful for malaria vaccine
PT prodn. and diagnosis
PS Claim 1; Fig 2A-1; 69pp; French.
CC This sequence corresponds to a Plasmodium falciparum strain K1
CC pre-erythrocytic liver stage antigen-3 (LSA-3) protein. The encoding
CC gene sequence was isolated by screening a P. falciparum strain 19/96
CC library with serum from a missionary treated by prophylaxis (for strain
CC screen a library generated from Thai strain K1. One clone contained a
CC 6.85 kb insert including the genomic sequence T78867. The gene comprises
CC a 1.8 kb region encoding 3 major blocks of tetrapeptide repeats
CC (especially the amino acid sequence VEES, VEEN, VEER, VAPS, VAPT, etc)
CC and a 3' hydrophobic region corresponding to a glycosyl-phosphatidyl-
CC inositol membrane anchoring sequence. The invention relates to new
CC polypeptides of at least 10 amino acids derived from the LSA-3 protein
CC with the exception of the peptides W24791-4. The LSA-3 peptides can be
CC used to raise antibodies and as vaccines for immunotherapy of malaria.
SQ Sequence 1786 AA;

Query Match 7.6%; Score 84.5; DB 1; Length 1786;
Best Local Similarity 21.6%; Pred. No. 14;
Matches 52; Conservative 45; Mismatches 79; Indels 65; Gaps 10;

Qy 27 SQPSPEVTAVALKLDLDRSTFRKLLKLVGALHGKDCREAVEQLGASANISEERLAVLL 86
Db 1183 SDSKEET---ESIKDKED-----VSLVVEVQDNMDSEVKVLELKNMEE----- 1229
Qy 87 AGTHLLQOALRLPPASLKPDAFOBELQELG-IPDGLIGDLASL-----AFG 132
Db 1229 -----LMKDAVEINDITSK---LIEETQELNEVEADLIKMEKLEKALSSEDSKEIID 1280
Qy 133 SORPLDSVAQQQ---GSSLPHVSVFRWRVDVAISTSAQSRSLOPSVLMOLK----- 182
Db 1281 AKDDTLERKVEEHDITTTDEVVELKDVDEEKTEKVSDDLKDLEEDILKEVKEKELESE 1340
Qy 182 -----LTDGSAHFVEVP---IAKF-----QELRYVALVKMAELEKCKERKLQ 223
Db 1341 ILEDYKELKTIETDILEEKEIEKDHFKFEAEAEIKDLEADILKEVSSLEVEEERKLE 1400
Qy 224 D 224
Db 1401 E 1401

RESULT 7
ID W53924
AC W53924
DT 20-AUG-1998 (first entry)
DE Pseudomonas pseudoalcaligenes lipase variant.
KW Lipase; variant; improved wash performance; removal; lipid stain;
KW reduced calcium dependence; one-cycle wash efficiency; detergent;
KW cleaning composition.
OS Synthetic.

OS Pseudomonas pseudoalcaligenes.
FH Key Location/Qualifiers
FT Misc_difference 216
FT /label= V216P
FT /note="Val optionally replaced with Pro"
FT Misc_difference 224
FT /label= S224T
FT /note="Ser optionally replaced with Thr"
FT Misc_difference 229
FT /label= D229G
FT /note="Asp optionally replaced with Gly"
FT Misc_difference 231
FT /label= P231A
FT /note="Pro optionally replaced with Ala"
FT Misc_difference 242
FT /label= H242R
FT /note="His optionally replaced with Arg"
FT Misc_difference 245
FT /label= K245M
FT /note="Lys optionally replaced with Met"
FT Misc_difference 250
FT /label= D250N
FT /note="Asp optionally replaced with Asn"
FT Misc_difference 268
FT /label= L268I
FT /note="Leu optionally replaced with Ile"
FT Misc_difference 272
FT /label= D272S
FT /note="Asp optionally replaced with Ser"
FT Misc_difference 275
FT /label= T275S
FT /note="Thr optionally replaced with Ser"
FT Misc_difference 286
FT /label= L286X
FT /note="Leu optionally replaced with Ile or Asn"
PN W09808939-A1.
PD 05-MAR-1998.
PF 26-AUG-1997; DK0345.
PR 29-AUG-1996; US-029190.
PR 27-AUG-1996; DK-000902.
PA (NOVO) NOVO-NORDISK AS.
PI Okkels JS, Svendsen A;
DR WPI: 98-230359/20.
PT Variants of lipase from Pseudomonas containing specific amino acid
PT substitutions - deletions or additions, having improved wash
PT performance in detergent formulations
PS Claim 5a; Page -: 104pp; English.
CC The present sequence represents a variant of a wild-type Pseudomonas
CC pseudoalcaligenes lipase. The variant lipase has better wash
CC performance than the original lipase. It has improved removal of
CC lipid stains, reduced calcium dependence, better compatibility with
CC detergents or their components, increased hydrophobicity, altered
CC substrate specificity and better one-cycle wash efficiency. The variant
CC lipase is used in detergent and cleaning compositions.
CC note: this sequence does not appear in the specification; it was created
CC using information provided.
SQ Sequence 289 AA;

Query Match 7.4%; Score 82.5; DB 1; Length 289;
Best Local Similarity 26.3%; Pred. No. 1.9;
Matches 49; Conservative 23; Mismatches 75; Indels 39; Gaps 9;

Qy 15 ADSHSGRVSLFG-SQSPSEVTAVALKLDLDRSTFRKLLKLVGALH-GKDCREAVEQLG 72
Db 74 AISGKGVNLVGHSHGPTVRYVAARPDIVASV-----TSVGAPHKGSDDTADFIRQI- 127
Qy 73 ASANLSEERLAVLLAGTHLLQOALRLPPASLKPDAFOBELQELGIPQDLIGLQSL--- 130
Db 127 PPSGAGEIVAGIVNGLIGALINFLSGSSPS-----PQNALGALESUKSE 171
Qy 130 ---AFGSRP---LLDSVAQQQSSSLPHVSVFRWRVDVAISTSAQSRSLOPSVLM---OLKL 182

Db 172 GAAAFNAKYPOGIPTACGSGAYKXVNGSVYSWS-----GTSPLTNVLPDSDLLGATSL 226

QY 183 TDGSAH 188
| | |

Db 227 TFGEAN 232

RESULT 8

W46270 ID W46270 standard; Protein; 758 AA.

AC W46270;

DT 17-AUG-1998 (first entry)

DE Moraxella catarrhalis strain 012E CopB outer membrane protein.

KW CopB gene; outer membrane protein; epitope; infection;

KW diagnosis; therapy; vaccine.

OS Moraxella catarrhalis strain 012E.

FH Key Location/Qualifiers

FT Region 10..270

FT /note="approximate location of conserved region"

FT 275..302

FT /note="approximate location of non-conserved

FT region 1"

FT 322..329

FT /note="approximate location of non-conserved

FT region 2"

FT 348

FT /note="approximate location of non-conserved

FT region 3"

FT 350..395

FT /note="approximate location of conserved region"

FT 400..416

FT /note="approximate location of non-conserved

FT region 4"

FT 437..450

FT /note="approximate location of non-conserved

FT region 5"

FT 455..480

FT /note="approximate location of conserved region"

FT 485..486

FT /note="approximate location of non-conserved

FT region 6"

FT 490..520

FT /note="approximate location of conserved region"

FT 526..541

FT /note="approximate location of non-conserved

FT region 7"

FT 545..580

FT /note="approximate location of conserved region"

FT 583..596

FT /note="approximate location of non-conserved

FT region 8"

FT 600..635

FT /note="approximate location of conserved region"

FT 637..640

FT /note="approximate location of non-conserved

FT region 9"

FT 645..680

FT /note="approximate location of conserved region"

FT 683..695

FT /note="approximate location of non-conserved

FT region 10"

FN W0906851-A2.

PD 19-FEB-1998.

PF 12-AUG-1997; U14221.

PR 12-AUG-1996; US-023832.

PA (TEXA.) UNIV TEXAS SYSTEM.

PI Aebi C, Cope LD, Hansen EJ;

DR WPI; 98-159542/14.

DR N-PSDB; V26360.

PT New Isolated Moraxella catarrhalis peptide(s) - which define

PT epitopes of the outer membrane protein used to develop products for

PT the diagnosis, prophylaxis and treatment of infection

PS Example 1; Page 91-94; 132pp; English.

CC This protein comprises the CopB outer membrane protein of Moraxella
catarrhalis strain 012E, encoded by an isolated copB gene (see
V26360). CopB represents an important antigenic determinant of M.
catarrhalis, and a specific epitope (see W46250-51) has been
identified in the non-conserved region 1 of strains 012E and
035E that is bound by monoclonal antibody 10f3. Claimed peptides
(see W46274-81) of this region can be used in the diagnosis,
treatment and prophylaxis (as vaccines) of M. catarrhalis
infection.

QY 16 DSHSGRVSPGLGSPSPPE-----VTAVAOALLKDLDRS-----TFRKLLKL----- 55

Db 295 NKYAGKGYLGSK-TPDGKKYVVDANGKLVDLDRNPTQRETYQKLNLEWTGKNLGF 353

QY 55 -----VVGALHGKDC-----REAVPOL-----GASANLSEERL 82

Db 354 ANEVTANVYKLEHGRNSSDQNTYITKDPKEIIDNVTPSNMHWVATGANINFDEKN 413

QY 83 AVLLAG---THTLLQALR-----LPPASLKPDAFQELQELGIPQDLIGDL 126

Db 414 HSPLKGFVGDHTLLKYGINVYHQKAVPPRSRKPVGIVHQEKTDAGIYLEAVNQI 466

RESULT 9

R75189 ID R75189 standard; Protein; 1196 AA.

AC R75189;

DT 30-MAY-1996 (first entry)

DE Osteoinductive retrovirus RFB-14 pol gene product.

KW RFB retrovirus; gag; pol; env; osteogenesis; osteoinductive protein;

KW bone development; osteoporosis; gene therapy; polymerase;

KW reverse transcriptase.

OS Retrovirus RFB-14.

PN DE4411718-Al.

PD 12-OCT-1995.

PF 05-APR-1994; 411718.

PR 05-APR-1994; DE-411718. 11

PA (GSFU-) GSF FORSCHUNGSZENTRUM UMWELT & GESUNDHEIT.

PI Erfile V, Gimbel W, Oestergaard M, Pedersen FS, Pedersen L;

PI Schmidt J, Strauss P;

DR WPI; 95-352078/46.

DR N-PSDB; Q94266.

PT RFB-14 retrovirus genome - and prodn. of osteo-inductive proteins

PS Claim 14; Fig 1; 46pp; German.

CC The full-length proviral genomic sequence of retrovirus RFB-14 has
been determined. The virus codes for an osteoinductive protein,
although the precise location of the coding region has not yet been
identified. The virus may be useful in gene therapy of bone growth
disorders such as osteoporosis. The present sequence is that of the
viral pol gene product.

QY 33 VTAVAOALLKDLDRSTFRKLLKLVVVGALHGKDCREAVPOLGASANLSEERL-----AVLLAG 88

Db 533 VAAIAVLTNDAGKLTMGQ---PLVILAPHAEE---EALVKQPPDRWLNSARNTHVQAMLLDT 587

QY 89 THTLLQALRLPPASLKPDAFQELQELGIPQDLIGDLASLAFSGQRPLDSV----- 142

Db 588 DRVQFGPVVALNPATLLP-----LPEEGAPHCLELAE-THGTRPDLTDQIPDADHT 640

QY 142 -----AQOQGSSLPVHSYFNRVVDVAISTASRSRLOPSVLMLKLTDSAGR 189

Query Match 7.4%; Score 82.5; DB 1; Length 1196;
Best Local Similarity 24.2%; Pred. No. 13;
Matches 54; Conservative 31; Mismatches 83; Indels 55; Gaps 10;


```

QY 82 LAVLLAG----- 89
D 229 LQALLFGDHRCTFRTMTALLPRSEPAFLPAHGQDTPVPPPPRPSAELEBSPPSADP 288
QY 89 ---THTLQALRLPPA-----SLKPDFAQELQELGIPQDLI--GDASL--AFGSQ 134
D 289 FLETILRLVRLVPPARASAPRLADPDAL-----AGFPQGLVNLSDPAALERLLDGE 342
QY 135 RPLL-----DSVAQOQGSSLPVSVYFRWRVDVAI--STSAQSRSL 172
D 343 EPLLLLLRPTAATGDPAPLHDPTSPATALLARRVAALQAAAEELRSL 392

RESULT 13
ID R76501 standard; Protein; 560 AA.
AC R76501;
DT 19-DEC-1995 (first entry)
DE Human MIS protein.
KW Mullerian inhibiting substance; MIS; ovary cancer; therapy.
OS Homo sapiens.
FH Key
FT 25..536
FT /label= Mat_protein
FT /note= "Claim 1, column 26"
PN US5427780-A.
PD 27-JUN-1995.
PF 30-OCT-1985; 792880.
PR 30-OCT-1985; US-792880.
PR 28-OCT-1986; US-923879.
PR 25-APR-1991; US-693764.
PR 05-OCT-1992; US-957061.
PA (BIOGEN ) BIOGEN INC.
PA (GEO ) GEN HOSPITAL CORP.
PI Cate RL, Donahoe PK;
DR WPI: 95-239996/31.
DR N-PSDB: Q92785.
DR Compns. comprising human Mullerian Inhibiting substance - useful for
PT treating cancer, esp. ovarian cancer.
PS Disclosure: Fig.6a-6k; 41pp; English.
CC The insert of bovine MIS cDNA clone p521 was used to isolate the
CC human MIS gene from a cosmid library. The DNA can be operatively
CC linked to expression control sequences and used in mammalian or
CC other host cells to produce MIS-like polypeptides.
SQ Sequence 560 AA;

Query Match 7.4%; Score 82; DB 1; Length 560;
Best Local Similarity 23.5%; Pred. No. 5.1;
Matches 54; Conservative 17; Mismatches 67; Indels 92; Gaps 9;

QY 29 PSPEVTAVALLL---KDLDRSTFRKLLKLVV---GALHGKDCREAVEQLGASANLSEER 81
D 170 PGPEVTVTRAGLPGAQSLCPSRDRYLVAVDRPAGAWRGSLALTLPGRGDSRLSTAR 229
QY 82 LAVLLAG----- 89
D 230 LQALLFGDHRCTFRTMTALLPRSEPAFLPAHGQDTPVPPPPRPSAELEBSPPSADP 289
QY 89 ---THTLQALRLPPA-----SLKPDFAQELQELGIPQDLI--GDASL--AFGSQ 134
D 290 FLETILRLVRLVPPARASAPRLADPDAL-----AGFPQGLVNLSDPAALERLLDGE 343
QY 135 RPLL-----DSVAQOQGSSLPVSVYFRWRVDVAI--STSAQSRSL 172
D 344 EPLLLLLRPTAATGDPAPLHDPTSPATALLARRVAALQAAAEELRSL 393

RESULT 14
R34712
ID R34712 standard; Protein; 3588 AA.
AC R34712;
DT 17-AUG-1993 (first entry)

```

```

DE Bacillus subtilis srfa operon ORF1 prod.
KW Multienzyme complex; surfactin synthetase; MCSS; ORF; surfactant.
OS Bacillus subtilis.
FH Key
FT 139..145
FT /note= "repeat sequence"
FT 1184..1190
FT /note= "repeat sequence"
FT 2223..2229
FT /note= "repeat sequence"
FT 3255..3261
FT /note= "repeat sequence"
FT 606..616
FT /note= "repeat sequence"
FT 1653..1663
FT /note= "repeat sequence"
FT 2698..2709
FT /note= "repeat sequence"
FT 858..880
FT /note= "repeat sequence"
FT 1897..1919
FT /note= "repeat sequence"
FT 2930..2949
FT /note= "repeat sequence"
PN EP-540074-A.
PD 05-MAY-1993.
PF 03-OCT-1992; 203037.
PR 09-OCT-1991; IT-MI2683.
PR 02-SEP-1992; IT-MI2044.
PA (ENIE ) ENIRICERCH SPA.
PI Carrera P, Cosmina P, De FERRA F, Grandi G, Perego M;
PI Rodriguez F;
DR WPI: 93-145447/18.
DR N-PSDB: Q40706.
DR Multi-enzyme complex surfactin synthetase DNA - is isolated from
PT Bacillus subtilis, and used for prodn. of surfactin for use in
PT pharmaceutical prods.
PS Claim 15; Page 27-34; 70pp; English.
CC The Bacillus subtilis chromosomal DNA region comprises the srfa
CC operon which encodes the multienzyme complex surfactin synthetase
CC (MCSS). Analysis of the sequence showed four regions potentially
CC coding for proteins, a zone upstream of the first ORF contg. the
CC srfa operon promoter and a presumed terminator positioned downstream
CC of the stop codon of the fourth ORF. ORF1 encodes a protein
CC (shown) which can be divided into 3 adjacent repeat regions with an
CC internal homology of ca. 1000 bases, more marked in the second half of
CC these regions. At the C terminal end of module 3 there is a region
CC (module A) of 500 amino acids which has no homology with modules 1-3.
CC Although ORF1 codes for a protein of unknown function, regions of the
CC protein are found to be highly homologous to synthetases coded by TycA
CC and GrsA (tyrocidin and gramicidin synthetase subunits I) and with
CC aminoadipyl cystein valine synthetase.
CC See also R34713-21.
SQ Sequence 3588 AA;

```

```

Query Match 7.3%; Score 81; DB 1; Length 3588;
Best Local Similarity 20.7%; Pred. No. 80;
Matches 56; Conservative 36; Mismatches 87; Indels 92; Gaps 11;

QY 1 MSALGRAAPVILHPAD-----SHSGRVSVFLGSPSPSEVTAVALLLKDLDRSTFRKLL 52
D 2279 LKEIDGELPVLTPDYSRPAVQTFEGDRIAF--SLEAAKADALRLAKETDSTLYMVL 2336
QY 53 -----KLVVGA-LHGKDCREAVEQLGASANLSEERLAVLLAGTHTTLQAL 97
D 2337 ASYSAFLSKTCGGDDIIVGSPVAGRSQADVSRVGMFVN-----TL---AL 2379
QY 98 RLPPASLKPDA-FOEELQELGIP-----ODLIGDLASLAFGSGORPLLDSDVAQOQGS 147
D 2380 RYTPKGEKTFADYLVNEVKETALSFAHQADYPLEDLIGNVQVQRTSSNPLFDVAFSNQNA 2439
QY 148 SLPHVSVFRWRVDVAISTSAQSRSLQPSVLMLQKLKLDGSAHREVFPIAKFQ----- 199

```

[illegible]

"

"

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 15, 2000, 01:46:14 ; Search time 35.18 Seconds
(without alignments)
92.094 Million cell updates/sec

Title: US-09-223-796-2

Perfect score: 1112

Sequence: 1 MSALGAAAPYLHPADSHG.....ALVLKEMAELEKKCKERKLQD 224

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 143561 seqs, 14463640 residues

Total number of hits satisfying chosen parameters: 143561

Minimum DB seq length: 0

Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database : Issued_Patents_AA.*

1: /cgn2_6/ptodata/2/iaa/5A_COMB.pep.*

2: /cgn2_6/ptodata/2/iaa/5B_COMB.pep.*

3: /cgn2_6/ptodata/2/iaa/6_COMB.pep.*

4: /cgn2_6/ptodata/2/iaa/PCTUS_COMB.pep.*

5: /cgn2_6/ptodata/2/iaa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	86.5	7.8	195	2	US-08-822-260-1
2	82.5	7.4	195	2	US-08-822-260-3
3	82	7.4	560	1	US-07-683-957B-1
4	79.5	7.1	681	1	US-08-083-590A-18
5	79.5	7.1	681	2	US-08-346-128-37
6	79.5	7.1	1078	1	US-08-264-534-32
7	79.5	7.1	1078	1	US-08-083-590A-11
8	79.5	7.1	1078	1	US-08-465-500-32
9	79.5	7.1	1078	2	US-08-346-128-32
10	79.5	7.1	2556	1	US-08-083-590A-20
11	78.5	7.1	363	1	US-08-530-950-6
12	78.5	7.1	399	1	US-08-530-950-10
13	78.5	7.1	399	2	US-08-874-186-92
14	77	6.9	3567	2	US-07-642-734C-4
15	76.5	6.9	393	1	US-08-530-950-8
16	76	6.8	689	1	US-08-221-817-18
17	76	6.8	689	1	US-08-454-439-18
18	76	6.8	689	4	PCT-US94-10487-18
19	75.5	6.8	226	3	US-08-966-318-1
20	75.5	6.8	226	3	US-08-966-318-1
21	75	6.7	635	2	US-08-484-101B-36
22	75	6.7	635	2	US-08-484-101B-50
23	74.5	6.7	334	5	5290690-10
24	74.5	6.7	335	5	5290690-9
25	74.5	6.7	1079	2	US-08-929-967-8
26	74	6.7	688	1	US-08-221-817-19
27	74	6.7	688	1	US-08-454-439-19
28	74	6.7	688	4	PCT-US94-10487-19
29	74	6.7	1111	1	US-08-317-450B-15

30 74 6.7 1163 1 US-08-173-497-4 Sequence 4, Appli
31 74 6.7 1163 1 US-08-286-889-4 Sequence 4, Appli
32 74 6.7 1163 1 US-08-485-618-4 Sequence 4, Appli
33 74 6.7 1163 1 US-08-362-652-4 Sequence 4, Appli
34 74 6.7 1163 2 US-08-605-672-4 Sequence 4, Appli
35 74 6.7 1163 2 US-08-482-293A-4 Sequence 4, Appli
36 74 6.7 1163 2 US-08-943-363-4 Sequence 4, Appli
37 74 6.7 1163 2 US-08-476-062A-44 Sequence 4, Appli
38 74 6.7 1163 4 PCT-US96-01314-44 Sequence 44, Appli
39 74 6.7 1193 1 US-08-317-450B-13 Sequence 13, Appli
40 74 6.7 3491 2 US-07-642-734C-2 Sequence 2, Appli
41 73.5 6.6 668 1 US-08-205-018-2 Sequence 2, Appli
42 73.5 6.6 859 1 US-08-395-580-2 Sequence 2, Appli
43 73.5 6.6 859 4 PCT-US95-02792-2 Sequence 2, Appli
44 72.5 6.5 532 1 US-08-188-228-44 Sequence 44, Appli
45 72.5 6.5 532 1 US-08-332-638-44 Sequence 44, Appli

ALIGNMENTS

RESULT 1
US-08-822-260-1
; Sequence 1, Application US/08822260
; Patent No. 5830660
; GENERAL INFORMATION:
; APPLICANT: Hillman, Jennifer L.
; APPLICANT: Goli, Surya K.
; TITLE OF INVENTION: NOVEL TUMORIGENESIS PROTEIN
; NUMBER OF SEQUENCES: 3
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/822,260
; FILING DATE: Herewith
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Billings, Lucy J.
; REGISTRATION NUMBER: 36,749
; REFERENCE/DOCKET NUMBER: PF-0247 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-855-0555
; TELEFAX: 415-845-4166
; TELEX:
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 195 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: UTRSN02
; CLONE: 2267574
; US-08-822-260-1

Query Match 7.8%; Score 86.5; DB 2; Length 195;
Best Local Similarity 22.1%; Pred. No. 0.067;
Matches 46; Conservative 33; Mismatches 82; Indels 47; Gaps 8;

CITY: New York
STATE: New York
COUNTRY: U.S.A.
ZIP: 10036
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/346.128
FILING DATE:
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/879,038
FILING DATE: 30-APR-1992
ATTORNEY/AGENT INFORMATION:
NAME: Mistrock, S. Leslie
REGISTRATION NUMBER: 18,872
REFERENCE/DOCKET NUMBER: 7326-009
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212 790-9090
TELEFAX: 212 8698864/9741
TELEX: 66141 PENNIE
INFORMATION FOR SEQ ID NO: 37:
SEQUENCE CHARACTERISTICS:
LENGTH: 681 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: unknown
MOLECULE TYPE: peptide
US-08-346-128-37

Query Match 7.1%; Score 79.5; DB 2; Length 681;
Best Local Similarity 28.9%; Pred. No. 2.7;
Matches 44; Conservative 17; Mismatches 76; Indels 15; Gaps 6;

QY 55 VVGLHGKDCREAEVGLGASANLSEERLAVL--LAGTHTLLOQALRLPPASLKPDFAQEE 112
DB 486 MVGLPHLSLAASALSQMSVQGLPSTRLATPHLVQTQQVQPNLQMQQQNLQANTQQQ 545
QY 113 --LQ-----ELGIPQDITGLDASLAFGSQRPLLDVSAQQCGSSLPVSVYERWRVDVA 162
DB 546 QSLQPPPPPPQPHLGVSSAASGHLGR-SFLSGERSQADV-QPLGFPSSLAVHTILPQESPA 603
QY 163 ISTSAQRSRQPSVLMQLKLTGSAHFEVPI 194
DB 604 LPTSLPS-SLVPPVTAAQFLTTPSQHSYSPV 634

RESULT 6
US-08-264-534-32
Sequence 32, Application US/08264534
Patent No. 5648464
GENERAL INFORMATION:
APPLICANT: Artavanis-Tsakonas, Spyridon et al.
TITLE OF INVENTION: Human No. 5648464ch And Delta, Binding Domains
TITLE OF INVENTION: In Toporythmic Proteins, And Methods Based Thereon
NUMBER OF SEQUENCES: 34
CORRESPONDENCE ADDRESS:
ADDRESSEE: Pennie & Edmonds
STREET: 1155 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: U.S.A.
ZIP: 10036
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/264,534
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/695,189
FILING DATE: 03-MAY-1991
ATTORNEY/AGENT INFORMATION:
NAME: Misrock, S. Leslie
REGISTRATION NUMBER: 18,872
REFERENCE/DOCKET NUMBER: 7326-004
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212 790-9090
TELEFAX: 212 8698864/9741
TELEX: 66141 PENNIE
INFORMATION FOR SEQ ID NO: 32:
SEQUENCE CHARACTERISTICS:
LENGTH: 1078 amino acids
TYPE: amino acid
TOPOLOGY: unknown
MOLECULE TYPE: protein
US-08-264-534-32

Query Match 7.1%; Score 79.5; DB 1; Length 1078;
Best Local Similarity 28.9%; Pred. No. 5.5; Mismatches 17; Indels 15; Gaps 6;
Matches 44; Conservative 17;
QY 55 VVGALHGKDCREAVEQLGASANLSEERLAVL--LAGTHLLQQALRLPPASLKPDFAFOEE 112
DB 893 MVGPLHSSLAASALSOMSYQGLPSTRLATOPHLVOTQVQPNLQMQQNLPANIQQ 952
QY 113 --LQ-----ELGIPQDLIGDLASLAFGSRPLDLSVAQQGSSLPVSYFRWRVDVA 162
DB 953 QSLQPPPPPPPHLGVSSAASGHLGR-SFLSGEPQADV-QPLGSPSSLAHVHTLTPQESPA 1010
QY 163 ISTSAQSRSLQPSVLMQLKLTGSAHREVPPI 194
DB 1011 LPTSLPS-SLVPPVTAQAFLTPPSQHSYSSPV 1041

RESULT 7
US-08-083-590A-11
Sequence 11, Application US/08083590A
Patent No. 5786158
GENERAL INFORMATION:
APPLICANT: Artavanis-Tsakonas, S. et al.
TITLE OF INVENTION: Therapeutic And Diagnostic Methods
TITLE OF INVENTION: And Compositions Based On No. 5786158ch Proteins And
TITLE OF INVENTION: Nucleic Acids
NUMBER OF SEQUENCES: 21
CORRESPONDENCE ADDRESS:
ADDRESSEE: Pennie & Edmonds
STREET: 1155 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: U.S.A.
ZIP: 10036
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/083,590A
FILING DATE: 25-JUN-1993
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Misrock, S. Leslie
REGISTRATION NUMBER: 18,872
REFERENCE/DOCKET NUMBER: 7326-015
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212 790-9090
TELEFAX: 212 8698864/9741

TELEX: 66141 PENNIE
INFORMATION FOR SEQ ID NO: 11:
SEQUENCE CHARACTERISTICS:
LENGTH: 1078 amino acids
TYPE: amino acid
TOPOLOGY: unknown
MOLECULE TYPE: protein
US-08-083-590A-11

Query Match 7.1%; Score 79.5; DB 1; Length 1078;
Best Local Similarity 28.9%; Pred. No. 5.5; Mismatches 17; Indels 15; Gaps 6;
Matches 44; Conservative 17;
QY 55 VVGALHGKDCREAVEQLGASANLSEERLAVL--LAGTHLLQQALRLPPASLKPDFAFOEE 112
DB 893 MVGPLHSSLAASALSOMSYQGLPSTRLATOPHLVOTQVQPNLQMQQNLPANIQQ 952
QY 113 --LQ-----ELGIPQDLIGDLASLAFGSRPLDLSVAQQGSSLPVSYFRWRVDVA 162
DB 953 QSLQPPPPPPPHLGVSSAASGHLGR-SFLSGEPQADV-QPLGSPSSLAHVHTLTPQESPA 1010
QY 163 ISTSAQSRSLQPSVLMQLKLTGSAHREVPPI 194
DB 1011 LPTSLPS-SLVPPVTAQAFLTPPSQHSYSSPV 1041

RESULT 8
US-08-465-500-32
Sequence 32, Application US/08465500
Patent No. 5789195
GENERAL INFORMATION:
APPLICANT: Artavanis-Tsakonas, Spyridon
APPLICANT: Muskavitch, Marc A.T.
APPLICANT: Fehon, Richard G.
APPLICANT: Rebay, Ilaria
APPLICANT: Blumweller, Cristine M.
APPLICANT: Shepard, Scott B.
TITLE OF INVENTION: HUMAN NOTCH AND DELTA, BINDING DOMAINS
TITLE OF INVENTION: IN TOPOTHMIC PROTEINS, AND METHODS BASED THEREON
NUMBER OF SEQUENCES: 34
CORRESPONDENCE ADDRESS:
ADDRESSEE: PENNIE & EDMONDS
STREET: 1155 Avenue of the Americas
CITY: New York
STATE: NY
COUNTRY: USA
ZIP: 10036-2711
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/465,500
FILING DATE: 05-JUN-1995
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Misrock, S. Leslie
REGISTRATION NUMBER: 18,872
REFERENCE/DOCKET NUMBER: 7326-034
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 790-9090
TELEFAX: (212) 869-8864/9741
INFORMATION FOR SEQ ID NO: 32:
SEQUENCE CHARACTERISTICS:
LENGTH: 1078 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: unknown
MOLECULE TYPE: protein
US-08-465-500-32

```
Query Match 7.1%; Score 79.5; DB 1; Length 1078;
Best Local Similarity 28.9%; Pred. No. 5.5;
Matches 44; Conservative 17; Mismatches 76; Indels 15; Gaps 6;

QY 55 VVGALHGKDCREAVEQIGASANLSEERLAVL--LACTHTLLQALRLPPASLKPDFAQEE 112
:||||| :|:|:| :|:|:| :|:|:| :|:|:| :|:|:| :|:|:| :|:|:| :|:|:| :|:|:|
Db 893 MVGPLHSSLAASALSOMSYQGLPSTRLATQPHLVQTQVQPNLQMQOONLQPANIQOQ 952
:||||| :|:|:| :|:|:| :|:|:| :|:|:| :|:|:| :|:|:| :|:|:| :|:|:| :|:|:|
QY 113 --LQ-----ELGIPQDLIGLASLAFGSRPLDLSVAQOQGSLLPHYSYFRWRVDVA 162
:||||| :|:|:| :|:|:| :|:|:| :|:|:| :|:|:| :|:|:| :|:|:| :|:|:| :|:|:|
Db 953 QSLQPPPPPPQPHLGVSSAASGHLGR-SFLSGEPQADV-QPLGSSSLAVHTILPQESPA 1010
:||||| :|:|:| :|:|:| :|:|:| :|:|:| :|:|:| :|:|:| :|:|:| :|:|:| :|:|:|
QY 163 ISTSAQSRSLQPSVLMQLKLTGSAHRFEVPI 194
:||||| :|:|:| :|:|:| :|:|:| :|:|:| :|:|:| :|:|:| :|:|:| :|:|:| :|:|:|
Db 1011 LPTSLPS-SLVPPVTTAAQFLTPPSQHSYSSPV 1041
:||||| :|:|:| :|:|:| :|:|:| :|:|:| :|:|:| :|:|:| :|:|:| :|:|:| :|:|:|

RESULT 9
US-08-346-128-32
; Sequence 32, Application US/08346128
; Patent No. 5856441
; GENERAL INFORMATION:
; APPLICANT: Artavanis-Tsakonas, Spyridon et al.
; TITLE OF INVENTION: Human No. 5856441ch And Delta, Binding Domains
; TITLE OF INVENTION: In Toporythmic Proteins, And Methods Based Thereon
; NUMBER OF SEQUENCES: 37
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pennie & Edmonds
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/346,128
; FILING DATE:
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/879,038
; FILING DATE: 30-APR-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Mistrock, S. Leslie
; REGISTRATION NUMBER: 18,872
; REFERENCE/DOCKET NUMBER: 7326-009
; TELEPHONE: 212 790-9090
; TELEFAX: 212 8698864/9741
; INFORMATION FOR SEQ ID NO: 32:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1078 amino acids
; TYPE: amino acid
; TOPOLOGY: unknown
; MOLECULE TYPE: protein
US-08-346-128-32

Query Match 7.1%; Score 79.5; DB 2; Length 1078;
Best Local Similarity 28.9%; Pred. No. 5.5;
Matches 44; Conservative 17; Mismatches 76; Indels 15; Gaps 6;

QY 55 VVGALHGKDCREAVEQIGASANLSEERLAVL--LACTHTLLQALRLPPASLKPDFAQEE 112
:||||| :|:|:| :|:|:| :|:|:| :|:|:| :|:|:| :|:|:| :|:|:| :|:|:| :|:|:|
Db 893 MVGPLHSSLAASALSOMSYQGLPSTRLATQPHLVQTQVQPNLQMQOONLQPANIQOQ 952
:||||| :|:|:| :|:|:| :|:|:| :|:|:| :|:|:| :|:|:| :|:|:| :|:|:| :|:|:|
QY 113 --LQ-----ELGIPQDLIGLASLAFGSRPLDLSVAQOQGSLLPHYSYFRWRVDVA 162
:||||| :|:|:| :|:|:| :|:|:| :|:|:| :|:|:| :|:|:| :|:|:| :|:|:| :|:|:|
Db 953 QSLQPPPPPPQPHLGVSSAASGHLGR-SFLSGEPQADV-QPLGSSSLAVHTILPQESPA 1010
:||||| :|:|:| :|:|:| :|:|:| :|:|:| :|:|:| :|:|:| :|:|:| :|:|:| :|:|:|
QY 163 ISTSAQSRSLQPSVLMQLKLTGSAHRFEVPI 194
:||||| :|:|:| :|:|:| :|:|:| :|:|:| :|:|:| :|:|:| :|:|:| :|:|:| :|:|:|
Db 1011 LPTSLPS-SLVPPVTTAAQFLTPPSQHSYSSPV 1041
:||||| :|:|:| :|:|:| :|:|:| :|:|:| :|:|:| :|:|:| :|:|:| :|:|:| :|:|:|

RESULT 10
US-08-590A-20
; Sequence 20, Application US/08083590A
; Patent No. 5786158
; GENERAL INFORMATION:
; APPLICANT: Artavanis-Tsakonas, S. et al.
; TITLE OF INVENTION: Therapeutic And Diagnostic Methods
; TITLE OF INVENTION: And Compositions Based On No. 5786158ch Proteins And
; TITLE OF INVENTION: Nucleic Acids
; NUMBER OF SEQUENCES: 21
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pennie & Edmonds
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/083,590A
; FILING DATE: 25-JUN-1993
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Mistrock, S. Leslie
; REGISTRATION NUMBER: 18,872
; REFERENCE/DOCKET NUMBER: 7326-015
; TELEPHONE: 212 790-9090
; TELEFAX: 212 8698864/9741
; INFORMATION FOR SEQ ID NO: 20:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2556 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: unknown
; MOLECULE TYPE: peptide
US-08-083-590A-20

Query Match 7.1%; Score 79.5; DB 1; Length 2556;
Best Local Similarity 28.9%; Pred. No. 21;
Matches 44; Conservative 17; Mismatches 76; Indels 15; Gaps 6;

QY 55 VVGALHGKDCREAVEQIGASANLSEERLAVL--LACTHTLLQALRLPPASLKPDFAQEE 112
:||||| :|:|:| :|:|:| :|:|:| :|:|:| :|:|:| :|:|:| :|:|:| :|:|:| :|:|:|
Db 2347 MVGPLHSSLAASALSOMSYQGLPSTRLATQPHLVQTQVQPNLQMQOONLQPANIQOQ 2406
:||||| :|:|:| :|:|:| :|:|:| :|:|:| :|:|:| :|:|:| :|:|:| :|:|:| :|:|:|
QY 113 --LQ-----ELGIPQDLIGLASLAFGSRPLDLSVAQOQGSLLPHYSYFRWRVDVA 162
:||||| :|:|:| :|:|:| :|:|:| :|:|:| :|:|:| :|:|:| :|:|:| :|:|:| :|:|:|
Db 2407 QSLQPPPPPPQPHLGVSSAASGHLGR-SFLSGEPQADV-QPLGSSSLAVHTILPQESPA 2464
:||||| :|:|:| :|:|:| :|:|:| :|:|:| :|:|:| :|:|:| :|:|:| :|:|:| :|:|:|
QY 163 ISTSAQSRSLQPSVLMQLKLTGSAHRFEVPI 194
:||||| :|:|:| :|:|:| :|:|:| :|:|:| :|:|:| :|:|:| :|:|:| :|:|:| :|:|:|
Db 2465 LPTSLPS-SLVPPVTTAAQFLTPPSQHSYSSPV 2495
:||||| :|:|:| :|:|:| :|:|:| :|:|:| :|:|:| :|:|:| :|:|:| :|:|:| :|:|:|

RESULT 11
US-08-530-950-6
; Sequence 6, Application US/08530950
; Patent No. 5736381
; GENERAL INFORMATION:
```

```

; APPLICANT: Davis, Roger J.
; APPLICANT: Raingeaud, Joel
; APPLICANT: Gupta, Shashi
; APPLICANT: Derijard, Benoit
; TITLE OF INVENTION: CYTOKINE-, STRESS-, AND
; TITLE OF INVENTION: ONCOPROTEIN-ACTIVATED HUMAN PROTEIN KINASE
; NUMBER OF SEQUENCES: 16
; CORRESPONDENCE ADDRESS:
; ADDRESS: Fish & Richardson P.C.
; STREET: 225 Franklin Street
; CITY: Boston
; STATE: MA
; COUNTRY: USA
; ZIP: 02110-2804
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/530,950
; FILING DATE: 19-SEP-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Fasse, J. Peter
; REGISTRATION NUMBER: 32,983
; REFERENCE/DOCKET NUMBER: 07917/010001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617/542-5070
; TELEFAX: 617/542-8906
; TELEX: 200154
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 363 amino acids
; TYPE: amino acid
; STRANDEDNESS: not relevant
; TOPOLOGY: linear
; US-08-530-950-6

```

```

Query Match      7.1%; Score 78.5; DB 1; Length 363;
Best Local Similarity 20.9%; Pred. No. 1.3;
Matches 57; Conservative 41; Mismatches 94; Indels 81; Gaps 11;

QY 20 GRVSFLGSPSPVETAV-----AQLLKDL-----RSTFRKLLKLVVGGALHGK-D 63
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 78 GSVNKMVHKPSGQIMAVKRIRSTVDEKEQKQLMDLVVRRSSDCPYIVQYGFALFREGD 137
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 64 CREAVEQLGASAN-----LSEERLAVLLAGTHTLQQALRLPPASLK 105
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 138 CWCIMELMSTSFDFKFKYKYSVLDVPIEILGKITLATVKALNH--LKENLKIHRDIK 195
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 106 PD-----AFQEELOELGIPQDLIGDLASLAFSGORPL-----LDSVAQOQGSSL 150
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 196 PSNILLDRSGNIKLCDFGISGLQVLDSIAKTRDAGCRPYMAPERIDPSASRGQYDVRSDW 255
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 150 -----PHYSYFRW-----RVDVAISTSAQSRSLQPSVL--MOLKLTGD 185
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 256 SLGITLVELATGRFPYKPNVNSFDLTQVVKGDPPQLNSSEERFSPFINFVNLCTKD 315
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 186 SAHFEVPIAKFQELRYSVALLKEMAELEKCC 218
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 316 ESKR-----PKYKELLKHPFILMYEERAVEVAC 343
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

```

RESULT 12
US-08-530-950-10
; Sequence 10, Application US/08530950
; Patent No. 5736381
; GENERAL INFORMATION:
; APPLICANT: Davis, Roger J.
; APPLICANT: Raingeaud, Joel

```

```

; APPLICANT: Gupta, Shashi
; APPLICANT: Derijard, Benoit
; TITLE OF INVENTION: CYTOKINE-, STRESS-, AND
; TITLE OF INVENTION: ONCOPROTEIN-ACTIVATED HUMAN PROTEIN KINASE
; NUMBER OF SEQUENCES: 16
; CORRESPONDENCE ADDRESS:
; ADDRESS: Fish & Richardson P.C.
; STREET: 225 Franklin Street
; CITY: Boston
; STATE: MA
; COUNTRY: USA
; ZIP: 02110-2804
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/530,950
; FILING DATE: 19-SEP-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Fasse, J. Peter
; REGISTRATION NUMBER: 32,983
; REFERENCE/DOCKET NUMBER: 07917/010001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617/542-5070
; TELEFAX: 617/542-8906
; TELEX: 200154
; INFORMATION FOR SEQ ID NO: 10:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 399 amino acids
; TYPE: amino acid
; STRANDEDNESS: not relevant
; TOPOLOGY: linear
; US-08-530-950-10

```

```

Query Match      7.1%; Score 78.5; DB 1; Length 399;
Best Local Similarity 20.9%; Pred. No. 1.5;
Matches 57; Conservative 41; Mismatches 94; Indels 81; Gaps 11;

QY 20 GRVSFLGSPSPVETAV-----AQLLKDL-----RSTFRKLLKLVVGGALHGK-D 63
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 114 GSVNKMVHKPSGQIMAVKRIRSTVDEKEQKQLMDLVVRRSSDCPYIVQYGFALFREGD 173
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 64 CREAVEQLGASAN-----LSEERLAVLLAGTHTLQQALRLPPASLK 105
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 174 CWCIMELMSTSFDFKFKYKYSVLDVPIEILGKITLATVKALNH--LKENLKIHRDIK 231
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 106 PD-----AFQEELOELGIPQDLIGDLASLAFSGORPL-----LDSVAQOQGSSL 150
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 232 PSNILLDRSGNIKLCDFGISGLQVLDSIAKTRDAGCRPYMAPERIDPSASRGQYDVRSDW 291
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 150 -----PHYSYFRW-----RVDVAISTSAQSRSLQPSVL--MOLKLTGD 185
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 292 SLGITLVELATGRFPYKPNVNSFDLTQVVKGDPPQLNSSEERFSPFINFVNLCTKD 351
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 186 SAHFEVPIAKFQELRYSVALLKEMAELEKCC 218
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 352 ESKR-----PKYKELLKHPFILMYEERAVEVAC 379
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

```

RESULT 13
US-08-874-186-92
; Sequence 92, Application US/08874186
; Patent No. 5989885
; GENERAL INFORMATION:
; APPLICANT: Teng, David H-F.
; APPLICANT: Tavtigan, Sean V.
; APPLICANT: Perry III, William L.
; APPLICANT: Skolnick, Mark H.

```

```
;
; TITLE OF INVENTION: SPECIFIC MUTATIONS OF MAP KINASE KINASE
; TITLE OF INVENTION: 4 (MKK4) IN HUMAN TUMOR CELL LINES IDENTIFY IT AS A TUMOR
; TITLE OF INVENTION: SUPPRESSOR IN VARIOUS TYPES OF CANCER
; NUMBER OF SEQUENCES: 96
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Venable, Baetjer, Howard & Civiletti, LLP
; STREET: 1201 New York Avenue, N.W., Suite 1000
; CITY: Washington
; STATE: DC
; COUNTRY: U.S.A.
; ZIP: 20005
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/874,186
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/782,482
; FILING DATE: 10-JAN-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Saxe, Stephen A.
; REGISTRATION NUMBER: 38,609
; REFERENCE/DOCKET NUMBER: 24884-121392-01
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-962-4848
; TELEFAX: 202-962-8300
; INFORMATION FOR SEQ ID NO: 92:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 399 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-874-186-92

Query Match 7.1%; Score 78.5; DB 2; Length 399;
Best Local Similarity 20.9%; Pred. No. 1.5;
Matches 57; Conservative 41; Mismatches 94; Indels 81; Gaps 11;

QY 20 GRVSFLGQSPEVTAV-----AQLLKDD-----RSTFRKLLKLVGALHGK-D 63
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 114 GSNVKNVHKPSGQIMAVKIRISTVDEKEQQLLMDLDVNMRSDDCPYIVQFYGALPREGD 173
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 64 CREAVEOLGASAN-----LSEERLAVLLAGHTHLLQOALRLPASLK 105
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 174 CWCIMELMSTSDFKYKVVSVLDDVPEILGKITLATVKLNH--LKENLKIHRDIK 231
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 106 PD-----AFQEELQELGIPQDLIGDLASLAFGSQRPL-----LDSVAQOQGSLSL----- 150
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 232 PSNILLDRSGNKLCDFGISGQLVDSIAKTRDAGCRPYMAPERIDPSASRGQYDVRSDDW 291
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 150 -----PHVSVFRW-----RVDVAISTSAQSRLQPSVL--MQLKLTGD 185
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 292 SLGITLYELATGRFPYKKNVSFVDQLTVQVKGDPPLSNSEEREFSFSPFNFNLCITKD 351
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 186 SAHREVPVIAKFOELRYSVLVLKEMAELEKCC 218
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 352 ESKR-----PKYKELLKHPFILMYERAVEVAC 379
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 14
US-07-642-734C-4
; Sequence 4, Application US/07642734C
; Patent No. 5824513
; GENERAL INFORMATION:
; APPLICANT: Katz, L
; APPLICANT: Donadio, S
; APPLICANT: Mcalpine, J B
; TITLE OF INVENTION: Recombinant DNA Method for Producing
```

```
;
; TITLE OF INVENTION: Erythromycin Analogs
; NUMBER OF SEQUENCES: 27
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Edward H. Gorman
; STREET: Abbott Laboratories D377/AP6D-2 One Abbott
; CITY: Park Rd
; CITY: Abbott Park
; STATE: IL
; COUNTRY: US
; ZIP: 60064-3500
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/642,734C
; FILING DATE: 17-JAN-91
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Drackners, Andreas M
; REGISTRATION NUMBER: 32652
; REFERENCE/DOCKET NUMBER: 4952.US.01
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 708-937-9396
; TELEFAX: 708-938-2623
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3567 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-07-642-734C-4

Query Match 6.9%; Score 77; DB 2; Length 3567;
Best Local Similarity 25.3%; Pred. No. 66;
Matches 44; Conservative 23; Mismatches 45; Indels 62; Gaps 9;

QY 33 VTAVA-----QLLKDL-DRSTFRKLLKLVGALHGKDCREAVQOLGASANLSERLAVLL 86
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1136 VTGAASFVGDQLVRWLADRGARLVL---AGACPGDDLLAAVEAGAS-----AVVC 1184
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 87 AGHTLLQALRLPP-----ASLKPDFAQEEELQE---LGIPQDLIGD- 126
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1185 AQDAALREALGDPEYALVHAGTLNFGSISVAPFEAFETAAKTALLAVLDELVDGR 1244
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 126 -----LASLAFGSQRPLDLSVAQO---QGSSLPHYSVFRW 157
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1245 AVEREYCVSVAGIWGAGMAAAGS--AYLDALAEHHRARGRSCTSVAWTPW 1296
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 15
US-08-530-950-8
; Sequence 8, Application US/08530950
; Patent No. 5736381
; GENERAL INFORMATION:
; APPLICANT: Davis, Roger J.
; APPLICANT: Raingeaud, Joel
; APPLICANT: Gupta, Shashi
; APPLICANT: Derijard, Benoit
; TITLE OF INVENTION: CYTOKINE-, STRESS-, AND
; TITLE OF INVENTION: ONCOPROTEIN-ACTIVATED HUMAN PROTEIN KINASE
; TITLE OF INVENTION: KINASES
; NUMBER OF SEQUENCES: 16
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson P.C.
; CITY: 225 Franklin Street
; CITY: Boston
; STATE: MA USA
; COUNTRY: USA
; ZIP: 02110-2804
; COMPUTER READABLE FORM:
```

;; MEDIUM TYPE: Floppy disk
;; COMPUTER: IBM PC compatible
;; OPERATING SYSTEM: PC-DOS/MS-DOS
;; SOFTWARE: PatentIn Release #1.0, Version #1.30
;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: US/08/530,950
;; FILING DATE: 19-SEP-1995
;; CLASSIFICATION: 435
;; ATTORNEY/AGENT INFORMATION:
;; NAME: Fasse, J. Peter
;; REGISTRATION NUMBER: 32,983
;; REFERENCE/DOCKET NUMBER: 07917/010001
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: 617/542-5070
;; TELEFAX: 617/542-8906
;; TELEX: 200154
;; INFORMATION FOR SEQ ID NO: 8:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 393 amino acids
;; TYPE: amino acid
;; STRANDEDNESS: not relevant
;; TOPOLOGY: linear
US-08-530-950-8

Query Match 6.9%; Score 76.5; DB 1; Length 393;
Best Local Similarity 21.4%; Pred. No. 2.5;
Matches 58; Conservative 41; Mismatches 95; Indels 77; Gaps 12;
QY 20 GRVFLGSGPSPEVTAV-----AQLKDLDD---RSTFRKLLKLVVGVGALHGK-D 63
DB 108 GSVNKMVHKPSQIMAVKRIKIRSTVDEKEQKQLMDLDVVMRSSDCPYIVQFYGALFREGD 167
QY 64 CREAVEQQLGASAN-----LSEERLA-VLLAGTHL--LQOALRLPPASLKP 107
DB 168 CWICHELKSTFDKFKYKYVSVLDDVIPPEILGKITLATVKALMHLKENLKIHRDIKPS 227
QY 108 -----AFQEEQLQELGIPQDLIGDLASLAFGSQRPL-----LDSVAQQGSSL----- 150
DB 228 NILDRSGMIKLCDFGISGQLVDSIAKTRDAGCRPYMAPERIDFSASRQGYDVRSDVWSL 287
QY 150 -----PHVSYFRW-----RVDVAISTSAQSRSLQPSVL--MOLKLTGDSA 187
DB 288 GITLYELATGRFPYPKNSVFDQLTQVVKGDPPQLSNSEEREFSFINFVNLCLTKDES 347
QY 188 HRFVEPIAKFQELRYSVLVLKEMAELEKCC 218
DB 348 KR-----PKYKELLKHPFILMYEERAVEVAC 373

Search completed: May 15, 2000, 03:22:30
Job time: 5776 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model
Run on: May 15, 2000, 01:51:10 ; Search time 52.16 Seconds
(without alignments)
251.784 Million cell updates/sec

Title: US-09-223-796-2
Perfect score: 1112
Sequence: 1 MSALGAAAPYLHPADSHG.....ALVLKEMAELEKCKERKLD 224

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 168808 seqs, 58629743 residues

Total number of hits satisfying chosen parameters: 168808

Minimum DB seq length: 0
Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database : PIR_63.*
1: pir1.*
2: pir2.*
3: pir3.*
4: pir4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	104.5	9.4	676	S41022	hypothetical prote
2	93	8.4	486	D72007	CT863 hypothetical
3	91	8.2	584	I39710	cellulose biosynth
4	89.5	8.0	433	T19632	hypothetical prote
5	89	8.0	3187	JC5837	364K Golgi complex
6	84.5	7.6	673	T00328	hypothetical prote
7	84.5	7.6	674	T19780	hypothetical prote
8	84.5	7.6	1558	B71603	RESA-H3 antigen pF
9	84	7.6	303	T06981	low-molecularweig
10	84	7.6	304	T06505	glutenin low molec
11	84	7.6	661	A55877	tumor necrosis fac
12	84	7.6	790	S67803	probable membrane
13	84	7.6	1363	A55875	xanthine dehydroge
14	84	7.6	4151	T13734	groovin gene prote
15	83.5	7.5	336	T26189	hypothetical prote
16	83.5	7.5	1583	T14176	probable phosphati
17	83	7.5	200	S72612	RNA helicase II -
18	83	7.5	302	F69000	cobalamin biosynth
19	82.5	7.4	817	T03852	protein phosphatas
20	82.5	7.4	907	TG0193	G protein-coupled
21	82	7.4	560	WFIHUM	mullerian inhibiti
22	81.5	7.3	734	E72271	5-methyltetrahydro
23	81.5	7.3	1354	T13363	phosphoribosylform
24	81	7.3	459	T19347	hypothetical prote
25	81	7.3	1295	T24587	hypothetical prote
26	80.5	7.2	315	T64065	acetyl-CoA carboxy
27	80.5	7.2	1196	1 GNMVRV	poli polyprotein -
28	80.5	7.2	1325	S16129	dynelin-associated
29	80.5	7.2	1337	A53824	nuclear pore membr
30	80.5	7.2	1479	T17401	transcription regu

31 80.5 7.2 2491 2 A57036 talin - slime mold
32 80 7.2 368 2 T04861 hypothetical prote
33 80 7.2 998 2 A45580 transcription fact
34 80 7.2 1048 1 BVECSO exonuclease (EC 3.
35 80 7.2 1324 2 S22863 DNA-binding protei
36 80 7.2 2555 2 C69681 peptide synthetase
37 80 7.2 3588 2 I40485 surfactin syntheta
38 79.5 7.1 502 2 T35072 probable anthranil
39 79.5 7.1 694 2 C72761 hypothetical prote
40 79.5 7.1 843 1 GNVWK pol polyprotein -
41 79.5 7.1 1196 1 GNMVGV pol polyprotein -
42 79.5 7.1 1466 2 S65741 1-phosphatidylinos
43 79.5 7.1 1928 2 JS0610 beta-galactosidase
44 79.5 7.1 2555 2 A40043 notch protein homo
45 79 7.1 431 2 T28697 hypothetical prote

ALIGNMENTS

RESULT 1

S41022
hypothetical protein T07C4.9 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C>Date: 06-Jan-1995 #sequence_revision 06-Jan-1995 #text_change 09-Sep-1997
C:Accession: S41022
R:Berks, M.
submitted to the EMBL Data Library, January 1994
A:Reference number: S41014
A:Accession: S41022
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-676 <BER>
A:Cross-references: EMBL:Z29443; NID:g1067051; PID:g443836
C:Genetics: 69/3; 161/1; 208/3; 227/1; 357/1; 505/3
A:introns: 69/3; 161/1; 208/3; 227/1; 357/1; 505/3
C:Superfamily: annexin repeat homology
F:373-444/Domain: annexin repeat homology <AX1>
F:445-516/Domain: annexin repeat homology <AX6>
F:528-600/Domain: annexin repeat homology <AX3>
F:604-675/Domain: annexin repeat homology <AX4>

Query Match 9.4%; Score 104.5; DB 2; Length 676;
Best Local Similarity 25.7%; Pred. No. 0.88;
Matches 47; Conservative 34; Mismatches 53; Indels 49; Gaps 10;
QY 43 LDRSTFRKLLKLVVYGALHGKDCREAVEQVLGASANLSEERLAVL--LAGTHTLLQALRLP 100
Db 50 LNNKFRKLCELY-----KESR-----SHEMLSEEQLATLYDCATTSTRL----- 90
QY 101 PASLKPDARFEEQLQELIGIPQDLI--GDL--ASLAFS-----QRELLDSVAQ 143
Db 90 -----PNDFADQEDL---KNIIIRYELFKACHAINSTDFIQKSEDLKDEKVALERIVE 141
QY 144 QGSSLPVSYFRWRVVDVAISTSAQSRSLQPSVLMQLKLTGSAHRFEVPIAKFOELRY 202
Db 142 QKLAESAKNKEDIANVNVIVVANSYVAKSLRPVINIQMTPTVGGDTN-FEFDIDSFQFRQ 200
QY 203 SVA 205
Db 201 QLA 203

RESULT 2

D72007
CT863 hypothetical protein - Chlamydia pneumoniae (strain CWL029)
C:Species: Chlamydia pneumoniae
C>Date: 23-Apr-1999 #sequence_revision 23-Apr-1999 #text_change 08-Oct-1999
C:Accession: D72007
R:Kallman, S.; Mitchell, W.; Marathe, R.; Lammel, C.; Fan, J.; Olinger, L.; Grimwood, R.; Nature Genet. 21, 385-389, 1999
A:Title: Comparative genomes of Chlamydia pneumoniae and C. trachomatis.

A:Reference number: A72000; MUID:99206506
A:Accession: D72007
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-486 <ARN>
A:Cross-references: GB:AE001682; GB:AE001363; NID:g4377344; PIDN:AD19159.1; PID:g437735
A:Experimental source: strain CML029
C:Genetics:
A:Gene: Cpn1022

Query Match 8.4%; Score 93; DB 2; Length 486;
Best Local Similarity 21.9%; Pred. No. 4.8;
Matches 54; Conservative 35; Mismatches 100; Indels 58; Gaps 8;

QY 27 SOPSPEVTAVALKLDLSTFRK-----LLKLVVG-----ALHKDCREAVEQLGASAN 76
DB 2 NMPVPSAVPSANITLKEDSTVSTASGILKTATGEVLVSVCTALESSSTDALISLALGOI 61
QY 77 LSEERLAVLLAGTHLLQALRLPP-----ASLKPDAFO-EELQ 114
DB 62 ILATOQELLLOSTN--VHOLLFPEVVELEIQVVDLLVQLEHAETITSEPQETOTQSR 119
QY 115 ELGIPDLIGDLASLAFSGORPLLOSVAQOQSSLPVSYFR-----WRVDVAI 163
DB 120 EOTLPQSSSKOSALSRLKPEISDKOQALQTPKDSAVRKHSEAPSPETQARASLSQ 179
QY 164 STSAOSRSLOP-----SVLMQLKLDGSAHRFEVPIAKFOELRYSVVALVKENAELEK 216
DB 180 ASSSORSLPPOESAPERTLLEQQKASSFS-----PLSQFAERKQKALITSKSHELK 233
QY 217 KCEKRLQ 223
DB 234 ERDQDRQ 240

RESULT 3
I39710
cellulose biosynthesis protein celd - Agrobacterium tumefaciens
C:Species: Agrobacterium tumefaciens
C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 03-Dec-1999
A:Accession: I39710
R:Matthysse, A.G.; White, S.; Lightfoot, R.
J. Bacteriol. 177, 1069-1075, 1995
A:Title: Genes required for cellulose synthesis in Agrobacterium tumefaciens.
A:Reference number: I39709; MUID:95164506
A:Accession: I39710
A:Status: translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-584 <RES>
A:Cross-references: GB:L38609; NID:g710486; PIDN:AAC41431.1; PID:g710488
C:Comment: This protein is required for cellulose biosynthesis.
C:Genetics:
A:Gene: celd
C:Superfamily: Agrobacterium tumefaciens cellulose biosynthesis protein celd

Query Match 8.2%; Score 91; DB 1; Length 584;
Best Local Similarity 25.0%; Pred. No. 8.8;
Matches 53; Conservative 27; Mismatches 76; Indels 56; Gaps 10;

QY 22 VSFGLSQSPETAVALKLDLSTFRKLLKLVVG-----ALHKDCREAVE 69
DB 263 VLWFGDQP-----LGALANIIDROK-RSVLEFYTGRDEDWKTPSPGLVLHGHCIRRAIE 315
QY 70 Q-----LG-----ASANLSEERLAVLAGI-----HTLQALRLPP 101
DB 316 QGFKYDFLGNPEPKYFFGPEPKLSCTLFTRSGDNLGGTLHPRSVRFVFEQALKLYK 375
QY 102 ASLKPDFAFEELEIGIPDGLIG-----DLASLAFGS-----QRPLLDVAQOQSSLPVSYF 155
DB 376 SGKKEASTAFQMLSAAPDHLGAQFGLANLAFDRGEFRKPRSPFSASSQQTNP--SSF 433

QY 156 RWRVDVAI-STSAOSRSLOPSVLMQLKLTGDS 186
DB 434 GWRIGETLRCATLSRASEPSQVTNRVPPGS 465

RESULT 4
TI9632
hypothetical protein C32A3.3 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 29-Oct-1999
C:Accession: TI9632; TI9975
R:Thomas, K.
submitted to the EMBL Data Library, February 1995
A:Reference number: TI9154
A:Accession: TI9632
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-433 <WIL>
A:Cross-references: EMBL:Z48241; PIDN:CAA8286.1; GSPDB:GN00021; CESP:C32A3.3
A:Experimental source: clone C32A3
R:Burton, J.
submitted to the EMBL Data Library, November 1996
A:Reference number: TI9206
A:Accession: TI9975
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-433 <W12>
A:Cross-references: EMBL:Z81449; PIDN:CAB03766.1; GSPDB:GN00021; CESP:C32A3.3
A:Experimental source: clone C46F11
C:Genetics:
A:Gene: CESP:C32A3.3
A:Map position: 3
A:Introns: 61/3; 94/3; 191/1; 229/3; 298/3; 356/2

Query Match 8.0%; Score 89.5; DB 2; Length 433;
Best Local Similarity 23.6%; Pred. No. 7.9;
Matches 57; Conservative 39; Mismatches 95; Indels 51; Gaps 12;

QY 6 AAAPYLHPADSHSGRVSFLGSG-----PSPEVTAVLAQLKLDLSTFRKLLKLVVGL 59
DB 27 SAVTVLHPRKLLSAFVSAMAARAASPOSQKITVVD--VIDLAASIGNDFEKLIDN-- 83
QY 60 HGKDCREAVEQLGASANLSEERLAVLLAGTHLLQALRLPPA--SLKPDFAFQELQELG 117
DB 83 YGNECVRGIMPKVISA---LETLEAAGNDRENEIMRLSKAVERLEQEKHQRNQHLK 139
QY 118 IPQDL-----IGDLASLA---FGSQRPILDSVAQOQSSLPVSYFRWRVDVAIS 164
DB 140 FEELEQVEKTYRKDIDDLQOMVKSILVNNRLSTTV-----SSLPN-----HADSPVS 188
QY 165 TSAOSRSLOPSVLMQLKLTGDSAHREFVPI-----AKFOELRYSVVALVKEMAELE 215
DB 189 TSMREADLK--WLEELK--EMSSQORDEIKALQKVDVYTCQVENLQNSIEKLIRQNEEL 245
QY 216 KK 217
DB 246 RK 247

RESULT 5
JC5837
364K Golgi complex-associated protein - rat
C:Species: Rattus norvegicus (Norway rat)
C:Date: 05-Mar-1998 #sequence_revision 13-Mar-1998 #text_change 20-Aug-1999
C:Accession: JC5837
R:Toki, C.; Fujiwara, T.; Sohda, M.; Hong, H.S.; Misumi, Y.; Ikehara, Y.
Cell Struct. Funct. 22, 565-577, 1997
A:Title: Identification and characterization of rat 364-kDa Golgi-associated protein
A:Reference number: JC5837; MUID:98093490
A:Accession: JC5837
A:Status: nucleic acid sequence not shown
A:Molecule type: mRNA

A:Residues: 1-3187 <TK>
A:Cross-references: DBJ:D2543; NID:g516825; PIDN:BAA05026.1; PID:di005567; PID:g516826
C:Comment: This protein plays a role in the formation and maintenance of the characteris
C:Superfamily: giantin
F:49-549,624-1176,1238-1707,1763-3114/Domain: coiled-coil leucine zipper #status predict
F:3165-3187/Domain: membrane anchor #status predicted <AD>

Query Match 8.0%; Score 89; DB 2; Length 3187;
Best Local Similarity 21.8%; Pred. No. 1,1e+02;
Matches 49; Conservative 47; Mismatches 95; Indels 34; Gaps 8;
QY 11 LHHPADSHGRVFLGSGSPPEVTAVALKLDLDRSTFRKLLKLVVGAHKGDCREAVEQ 70
DB 1221 LHAQPSHGETATL-QATVSAQIQDOLKEIE--VEKEELEKISSTSELTKKSEV 1276
QY 71 LGASANLSEERLAVLLAGHTLLQALRPPASLKPDFAQEEL-QSLGIPQDLGLDLSL 129
DB 1277 LLLQEOINQGLEI-----ONLKAASHAKAHTQLKLESSQKIADLEHL 1324
QY 130 AFGSORPLDSVAQOQSSLPVSYF-----RWRVDVAISTSAQS-----RSLQPSVLM 178
DB 1325 K--TLQPELETLOKHVGQKEEVSVLVGQGEKEQTLTTVTQEMEQERLIKALHTQLEM 1382
QY 179 QLKLDGSAHREVPVTAQFQELRYVALVLEKAELEKKCEKRLQ 223
DB 1383 QAKEHEERUKQVQVCECLKQPKP-----LEESAKQQLQRLQ 1423

RESULT 6
T00328
hypothetical protein KIAA0552 - human
C:Species: Homo sapiens (man)
C:Date: 01-Feb-1999 #sequence_revision 01-Feb-1999 #text_change 08-Oct-1999
C:Accession: T00328
R:Nagase, T.; Ishikawa, K.; Miyajima, N.; Tanaka, A.; Kotani, H.; Nomura, N.; Ohara, O.
DNA Res. 5, 31-39, 1998
A:Title: Prediction of the coding sequences of unidentified human genes. IX. The complet
A:Reference number: 214086; MUID:98290545
A:Accession: T00328
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-673 <NA>
A:Cross-references: EMBL:AB011124; NID:d1185394; PIDN:BAA25478.1; PID:di026408
A:Experimental source: brain
C:Genetics:
A:Note: KIAA0552

Query Match 7.6%; Score 84.5; DB 2; Length 673;
Best Local Similarity 25.2%; Pred. No. 35;
Matches 54; Conservative 34; Mismatches 67; Indels 59; Gaps 12;
QY 15 ADSHGRVSLFG-----SOPSPVTAVALKLDLDRSTFRKLLKLVVGAH 60
DB 287 SSSMGRPHGLSGGGGLFPACSPSP-----SALIQEELRMEK--EQEVAAL- 339
QY 61 GKDCREAVEQLGAS-ANLSEER-----LAVLLAGHTLLQALRPPASLKPDFAQF- 112
DB 339 ----RRSLQSEAAVAQVLEERQKAWERLAELOGCSKQLQOVAR-----RAQRAQOQ 388
QY 112 -ELQELGIPQD--LIGDLASLAFSGSORPLDSVA---OOQSSLPVSYFVRV-----D 160
DB 389 LOLQRLQDQKQLQEEAARLMRQEELEDKVAACQEQADFLPRIETKWEVCQKAGE 448
QY 161 VAI-----SFSASRSRLOPSVLMQLKTDGSA 187
DB 449 ISLLKQQLKDSQADVSKLSEIVGLRSQRLREGRA 482

RESULT 7
T19780
hypothetical protein C36B1.9 - Caenorhabditis elegans

C:Species: Caenorhabditis elegans
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C:Accession: T19780
R:Lennard, N.
submitted to the EMBL Data Library, September 1996
A:Reference number: Z19176
A:Accession: T19780
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-674 <WIL>
A:Cross-references: EMBL:Z80215; PIDN:CAB02274.1; GSPDB:GN00019; CESP:C36B1.9
A:Experimental source: clone C36B1
C:Genetics:
A:Gene: CESP:C36B1.9
A:Map position: 1
A:Introns: 131/1; 179/1; 266/3; 468/2

Query Match 7.6%; Score 84.5; DB 2; Length 674;
Best Local Similarity 21.2%; Pred. No. 35;
Matches 39; Conservative 30; Mismatches 60; Indels 55; Gaps 7;
QY 63 DCREAVEQLGASANLSEERLAVLLAGHTLLQALRPPASLKPDFAQFQELG--IPQ 120
DB 111 ECKINEQVPATVPINFQIMQIL---TTLSLVKVLQTPPDQDEIPKY-ENFERLGTETP- 166
QY 121 DLIGDLASLAFSGSORPLDSVAQOQSSLPVSYFVRVDAISTSAQSRSLOPSVLMQL 180
DB 166 --IADLVNLTWGLYEHKAVFNA-----INLRGRKDKTSIRPSVV--- 205
QY 181 KLTGSAHREVPVIAKQFQELRYVALVLEKMA-----ELEKKCEK 220
DB 205 -----HYDSEFKKIDEVERSMNRVLHNVIKLSQGETLPRPAPTITIPYQLDREYR 257
QY 221 KIQD 224
DB 258 SLQD 261

RESULT 8
B71603
RESA-H3 antigen PFB0915w - malaria parasite (Plasmodium falciparum)
C:Species: Plasmodium falciparum
C:Date: 13-Nov-1998 #sequence_revision 13-Nov-1998 #text_change 07-May-1999
C:Accession: B71603
R:Portea, M.J.; Tettelin, H.; Carucci, D.J.; Cummings, L.M.; Aravind, L.; Koonin, E.;
Science 282, 1126-1132, 1998
A:Title: Chromosome-2 sequence of the human malaria parasite Plasmodium falciparum.
A:Reference number: A71600; MUID:99021743
A:Accession: B71603
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-1558 <GAR>
A:Cross-references: GB:AF001424; GB:AE001362; NID:g3845307; PID:g3845309; TIGR:PFB091
A:Experimental source: clone 3D7
C:Genetics:
A:Gene: PFB0915w

Query Match 7.6%; Score 84.5; DB 2; Length 1558;
Best Local Similarity 21.6%; Pred. No. 99;
Matches 52; Conservative 45; Mismatches 79; Indels 65; Gaps 10;
QY 27 SQSPSEVTAVALKLDLDRSTFRKLLKLVVGAHKGDCREAVEQLGASANLSEERLAVLL 86
DB 951 SSDSKEET---ESIKDKEK-----VSLVVEVQDNDMDSEVKVLELKNMEE----- 997
QY 87 AGHTLLQALRPPASLKPDFAQFQELG-IPQDLGLDLSL-----AFG 132
DB 997 -----LMKDAVEINDITSK-~~9~~LIEETOELNEVEADLTKMEKLEKALSSEDSKEIID 1048
QY 133 SQRPLLDSVAQOQ---GSSLPVSYFVRVDAISTSAQSRSLOPSVLMQLK----- 182

Db 1049 AKDPTLEKVEEEDHITTTLDVVELVDKEDKIERVSDLKOLEEDILKEYKEIKELESE 1108
QY 182 -----LTDGSAHRFPV---IAKF-----QELRYVALVLMKMAELEKCKERKIQ 223
Db 1109 ILEDYKELKTIETDILEKKEIEKDHFEEAEIKDLEADILKEVSSLEVEEKKLE 1168
QY 224 D 224
Db 1169 E 1169
RESULT 9
T06505
low-molecular-weight glutenin storage protein - wheat
C:Species: Triticum aestivum (common wheat)
C>Date: 30-Apr-1999 #sequence_revision 30-Apr-1999 #text_change 20-Aug-1999
C:Accession: T06505
R:Anderson, O.D.; Cassidy, B.; Dvorak, J.
submitted to the EMBL Data Library, January 1997
A:Reference number: Z15843
A:Accession: T06505
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-303 <AND>
A:Cross-references: EMBL:U86028; NID:g1857655; PIDN:AA84877.1; PID:g1857656
A:Experimental source: cv. Cheyenne
C:Genetics:
A:Map position: 1
A:Function:
A:Description: seed storage protein in the endosperm
C:Superfamily: gliadin
C:Keywords: seed; storage protein
Query Match 7.68; Score 84; DB 2; Length 303;
Best Local Similarity 27.38; Pred. No. 14;
Matches 38; Conservative 18; Mismatches 67; Indels 16; Gaps 4;
QY 99 LPPA---SLKPDFAQEELQELGIPQDLIGLASLAFGSQRPLDLS---VAQOQSSSLPHV 152
Db 38 LPPQSFSSQPPFSQQOQOPL--PQPSFSQQPPFSQQOQOPLSQPPFSQQOQOPLVLPQ 95
QY 153 SYFRWRVDVAISTSAQSRSL-----QPSVLMLQKLKTDGSAHREVEPIAKFQELRYSV 204
Db 96 SPFSQQOQLVLPQQOQOQLVQOQIPVQPSVLQQLNPKVFLQOQCSVPAMPQRLARSQ 155
QY 205 ALVLKEMAELEKCKERKIQ 223
Db 156 MWQOSSCHVMQOCCQQLQ 174
RESULT 10
T06505
glutenin low molecular weight chain precursor (B-I) - wheat
C:Species: Triticum aestivum (common wheat)
C>Date: 23-Apr-1999 #sequence_revision 23-Apr-1999 #text_change 12-Nov-1999
C:Accession: T06505
R:Okita, T.W.; Cheesbrough, V.; Reeves, C.D.
J. Biol. Chem. 260, 8203-8213, 1985
A:Title: Evolution and heterogeneity of the alpha/beta-type and gamma-type gliadin DNA
A:Reference number: A92541; MUID:85234522
A:Accession: T06505
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-304 <OKI>
A:Cross-references: EMBL:M1077; NID:g170729; PIDN:AAA34285.1; PID:g170730
C:Superfamily: gliadin
C:Keywords: seed; storage protein
F:1-23/Domain: signal sequence #status predicted <SIG>
F:24-304/Product: glutenin low molecular weight chain (B-I) #status predicted <MAT>

Query Match 7.68; Score 84; DB 2; Length 304;
Best Local Similarity 27.38; Pred. No. 14;
Matches 38; Conservative 18; Mismatches 67; Indels 16; Gaps 4;
QY 99 LPPA---SLKPDFAQEELQELGIPQDLIGLASLAFGSQRPLDLS---VAQOQSSSLPHV 152
Db 38 LPPQSFSSQPPFSQQOQOPL--PQPSFSQQPPFSQQOQOPLSQPPFSQQOQOPLVLPQ 95
QY 153 SYFRWRVDVAISTSAQSRSL-----QPSVLMLQKLKTDGSAHREVEPIAKFQELRYSV 204
Db 96 SPFSQQOQLVLPQQOQOQLVQOQIPVQPSVLQQLNPKVFLQOQCSVPAMPQRLARSQ 155
QY 205 ALVLKEMAELEKCKERKIQ 223
Db 156 MWQOSSCHVMQOCCQQLQ 174
RESULT 11
A55877
tumor necrosis factor type 1 receptor associated protein TRAP-1 - human
C:Species: Homo sapiens (man)
C>Date: 23-Mar-1995 #sequence_revision 23-Mar-1995 #text_change 05-Nov-1999
C:Accession: A55877
R:Song, H.Y.; Dunbar, J.D.; Zhang, Y.X.; Guo, D.; Donner, D.B.
J. Biol. Chem. 270, 3574-3581, 1995
A:Title: Identification of a protein with homology to hsp90 that binds the type 1 tum
A:Reference number: A55877; MUID:95181307
A:Accession: A55877
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-661 <SON>
A:Cross-references: GB:U12595; NID:g687236; PIDN:AAA87704.1; PID:g687237
C:Superfamily: heat shock protein 90
Query Match 7.68; Score 84; DB 2; Length 661;
Best Local Similarity 23.88; Pred. No. 37;
Matches 46; Conservative 39; Mismatches 76; Indels 32; Gaps 10;
QY 49 RKLLKVVGLHKGDCREA-VEOLGASANLSEERLAVLLAGTHTLLQQLRLPPASLKPD 107
Db 80 KKLDDIVARSLISE--KEVFIRELISNASDALEKL-----RHLVSDGQALPEMEIHLQ 131
QY 108 AFQEE---LQELGI---PQDLIGLASLAFGSQRPLDLSVAQOQSSSLPHVSY---FR 156
Db 132 TNAEKGTITTDGTGIGTQELVSNLGTIARSGSKAFDLALQNAEASSKIIGFGVGFY 191
QY 157 WRVDVAISTSAQSRSLQPSVLMLQKLKTDGSAHREVEPIAKFQELRYSVVALVKEMAELEK 216
Db 192 SAFMVADREVEYSRAAPGSLGYQLWLSGSG-VFE--IAEASGVRTGKIII---HLKS 244
QY 217 KC-----ERKIQD 224
Db 245 DCKEFSSEARVRD 257
RESULT 12
S67803
probable membrane protein YDL239c - yeast (Saccharomyces cerevisiae)
N:Alternate names: hypothetical protein D0771
C:Species: Saccharomyces cerevisiae
C>Date: 12-Jul-1996 #sequence_revision 12-Jul-1996 #text_change 29-Oct-1999
C:Accession: S67803
R:Alt-Moerbe, J.; Schneider, C.; Moro, M.
submitted to the Protein Sequence Database, July 1996
A:Reference number: S67798
A:Accession: S67803
A:Molecule type: DNA
A:Residues: 1-790 <ALT>
A:Cross-references: EMBL:Z74287; NID:g1431405; PID:e2533369; PID:g1431406; GSPDB:GN000
A:Experimental source: strain S288C
C:Genetics:
A:Gene: MIPS:YDL239c

A:Map position: 4L
C:Keywords: transmembrane protein
F:706-722/Domain: transmembrane #status predicted <TM>

Query Match 7.6%; Score 84; DB 2; Length 790;
Best Local Similarity 21.4%; Pred. No. 46;
Matches 57; Conservative 41; Mismatches 86; Indels 82; Gaps 12;

QY 11 LHPADSHSGRVSEFGSPSEVAVAGL-----LKDLSRSTFRKLLKLVVAGL-- 60
DB 517 LHFQQQSHKIKYIQKE---ALKEQQFRLEKRRWHIDILNKEFNQKLSKLGKIL 573
QY 60 -----HGKDCRAVEOLGASNL-----SERLAVLLAGHTL 92
DB 574 SEKIQKNAEDKLDYMNHEQVEIKLQNALIASRWSTQIOESNTHKKITDELAKGOS- 633
QY 93 LQALRLPPA---SLKPDFAQELQELGIPQDLIGDLASLAFSGSRPLDLSVAQQGSSLP 150
DB 633 --EILKLEETILSLKEDVQEKLN---LKKLYGD-----PSTELNFEVTKSFP 676
QY 151 HVS---YFRWRVDVAIS--TSAQSRSLQPSVLMQKLTDGSAHREVPYIAKEQ----- 199
DB 677 HITKEYDSGLDILDTDYVQSQNLKLLIVL-----DIPKTLKIVPTIVI 726
QY 199 ELRYSVLVLKEMAELEKKCKERKLQD 224
DB 727 QLRCELTLTKFANDLNLKVFQKQLD 752

RESULT 13
A55875
xanthine dehydrogenase (EC 1.1.1.204) - Emericella nidulans
N:Alternate names: purine hydroxylase I
C:Species: Emericella nidulans, Aspergillus nidulans
C:Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 10-Sep-1999
C:Accession: A55875; S49819
R:Glatigny, A.; Scazzocchio, C.
J. Biol. Chem. 270, 3534-3550, 1995
A:Title: Cloning and molecular characterization of hxa, the gene coding for the xanthine
A:Reference number: A55875; MUID:95181302
A:Molecule type: DNA
A:Residues: 1-1363 <GIA>
A:Cross-references: EMBL:X82827; NID:9577730; PIDN:CAA58034.1; PID:9577731
C:Genetics:
A:Gene: hxa

A:Introns: 83/3; 151/3; 803/2
C:Superfamily: xanthine dehydrogenase; ferredoxin (2Fe-2S) homology
C:Keywords: 2Fe-2S; FAD; flavoprotein; iron-sulfur protein; metalloprotein; molybdenum;
F:57-104/Domain: ferredoxin (2Fe-2S) homology <FER1>
F:73,78,81,103/Binding site: 2Fe-2S cluster (Cys) (covalent) #status predicted

Query Match 7.6%; Score 84; DB 1; Length 1363;
Best Local Similarity 22.5%; Pred. No. 92;
Matches 46; Conservative 35; Mismatches 81; Indels 42; Gaps 10;

QY 23 SFLGSPSEVAVAGLILK-DLDRST-FRKLKLVVAGLHGKDCREAVEQIGASNLSEE 80
DB 155 ALLRNDPKPSEHVAEAFDGNLCRGTGYRILDAASQTSPTGCGKARANGSGCCMEQ 214
QY 81 RLAVLLAGHTLLOALRLPPASLK-----PDFAQELQ-ELGIPQDL-IGDLASLAFGS 133
DB 215 K-----GTNGCKSGSSEETEDVKKHFPASDFIKYKPDTELIFPPSLWKHELRLPAGN 268
QY 134 QRLDLSVAQQGSSLPVSVFRWRVDVAISTSAQSRSLQPSVLMQKLTDGSAH-RFEV 192
DB 269 KRK-----KWRPVTVQQLLEIKSHHPDA-----KLIGSTETQIEI 305
QY 193 PIAKQELRYSVLVLKEMAELEK 216
DB 306 ---KFKOMRYGASVYGLDLAELRQ 326

RESULT 14

T13734
groovin gene protein - fruit fly (Drosophila melanogaster) (fragment)
C:Species: Drosophila melanogaster
C:Date: 13-Aug-1999 #sequence_revision 13-Aug-1999 #text_change 13-Aug-1999
C:Accession: T13734
R:Strumpf, D.; Volk, T.
J. Cell Biol. 143, 1259-1270, 1998
A:Title: Kakapo, a novel cytoskeletal-associated protein is essential for the restric
A:Reference number: Z17746
A:Accession: T13734
A:Status: preliminary; translated from GB/EMBL/DBDJ
A:Molecule type: mRNA
A:Residues: 1-4151 <STR>
A:Cross-references: EMBL:Y09430; NID:el372000; PID:el372001; PIDN:CAA70581.1
C:Genetics:
A:Gene: grv
A:Map position: 2

Query Match 7.6%; Score 84; DB 2; Length 4151;
Best Local Similarity 19.1%; Pred. No. 3.7e+02;
Matches 56; Conservative 51; Mismatches 108; Indels 78; Gaps 10;

QY 1 MSALGAAAPYLHHPADSHSGRVSEFG-----SQPS---PEVT----- 35
DB 2430 LGILEQALPLSEHFADSHQGLTAWDDMEQIISRLSPALRPDQITLQODKNERLLOSLA 2489
QY 35 -----AVALLKLDLDRSTFRKLL-----KLVVGALHGKDCREAVEQIGASA 75
DB 2490 EHKPILDKLNTGEALGALVADDGAKINEILDTDNARYAALRLERQALLESALQES 2549
QY 76 NLSEERLAVLAGHTLLOALRLPPASLKPDFAQELQELGIPQDLIGDLASLAFGS-Q 134
DB 2550 SQFSDKLEGLMLRALANTVDQVNLDPPLSALPQKIREQIEDNALMDDL-DKRDQAFSAVQ 2608
QY 135 RPLDLSVAQQGSSLPVSVFRWRV-----DVAISTSAQSRSL-----QP----- 175
DB 2609 RAANDVIATAKAGNKADPAVRDPAKLEKLNLMNDVQNAKTKRGSLLDDILSVAEFFWKQL 2668
QY 175 -SVLMQLKLTGDSAHREVPYIAK-----FQELRYSVLVLKEMAELEK 216
DB 2669 NSVVKTLKDLSEETLSQEPAPPAQPDIKKQVQLQELRHEIDQTKPEVEQYR 2721

RESULT 15

T26189
hypothetical protein W05B5.3 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C:Accession: T26189
R:McLay, K.
submitted to the EMBL Data Library, November 1996
A:Reference number: Z20167
A:Accession: T26189
A:Status: preliminary; translated from GB/EMBL/DBDJ
A:Molecule type: DNA
A:Residues: 1-336 <WIL>
A:Cross-references: EMBL:Z82071; NID:el298254; PIDN:CAB04919.1; GSPDB:GN00019; CESP:W
A:Experimental source: clone W05B5
C:Genetics:
A:Gene: CESP:W05B5.3
A:Map position: 1
A:Introns: 5/1; 59/3; 109/3; 233/3

Query Match 7.5%; Score 83.5; DB 2; Length 336;
Best Local Similarity 24.8%; Pred. No. 17;
Matches 51; Conservative 24; Mismatches 78; Indels 53; Gaps 8;
QY 6 AAAPYLHHPADSHSGRVSEFGSPSEVAVAGLILKLDLDRSTFRKLLK-----LVVAGLH 60

```
Db 140 SANPIIQAISARIQRPOTVPILTAIAVCQADLLPESQOPMLLAERLWCYLKLG 199
QY 61 GKDCREAVEQIGASANLSEERLAVLLAGTHTLLQ---QALR----- 99
Db 200 G-----IQSLATAPS-----LLADVTTLRQWHSDDLQMSQISQHFSONLLIAPVA 245
QY 99 -----LPPASLKPDAPQEELQELGIPQDLIGDLASLAFGSRP--LLDSVAQQQ-GSSL 149
Db 246 AAAPVLLPPAFLSPPASATSTSSSSVKSEFIERHPSTIASILERPRRISSSGAQEPNLNLSL 305
QY 150 PHVSYFRWRVDVAISTSAQSRLOPS 175
Db 306 PHV---RHQVKRDVDSDEOLEMKVS 328
```

Search completed: May 15, 2000, 03:23:35
Job time: 5545 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 15, 2000, 02:06:10 ; Search time 44.1 Seconds
(without alignments)
154.692 Million cell updates/sec

Title: US-09-223-796-2
Perfect score: 1112
Sequence: 1 MSALGAAAPYLHPADSHG.....ALVLKMAELEKCKERKIQD 224

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 83857 seqs, 30454973 residues

Total number of hits satisfying chosen parameters: 83857

Minimum DB seq length: 0

Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database : SwissProt_38.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	84.5	7.6	673	1 Y552_HUMAN	O60299 homo sapien
2	84.5	7.6	890	1 RB6K_HUMAN	O95235 homo sapien
3	84	7.6	304	1 GDB1_WHEAT	P04729 triticum ae
4	84	7.6	661	1 TRAL_HUMAN	Q12931 homo sapien
5	84	7.6	1363	1 XDHE_EMENI	Q12553 emericeella
6	83	7.5	302	1 COBK_METHH	O27083 methanobact
7	82	7.4	560	1 MIS_HUMAN	P03971 homo sapien
8	81.5	7.3	1354	1 PURA_DROME	P35421 drosophila
9	80.5	7.2	315	1 ACCA_HAEIN	P43872 haemophilus
10	80.5	7.2	1196	1 POL_MLVVD	P11227 radiation m
11	80.5	7.2	1280	1 DYNA_RAT	P28023 rattus norv
12	80.5	7.2	1337	1 P152_YEAST	P39685 saccharomyc
13	80.5	7.2	2491	1 TALA_DICDI	P54633 dictyosteli
14	80	7.2	314	1 IUNH_CRIFA	Q27346 crithidia f
15	80	7.2	887	1 RB6K_MOUSE	P97329 mus musculu
16	80	7.2	1048	1 SBCC_ECOLI	P13458 escherichia
17	80	7.2	2555	1 PPS3_BACSU	P39847 bacillus su
18	79.5	7.1	572	1 PRIM_STAOU	O05338 staphylococ
19	79.5	7.1	843	1 POL_MLVAK	P03357 akr murine
20	79.5	7.1	1196	1 POL_MLVAV	P03356 akr murine
21	79.5	7.1	1928	1 LPH_RAT	Q02401 rattus norv
22	79	7.1	1194	1 BCHA_RHOCA	P26162 rhodobacter
23	79	7.1	1270	1 DYNA_HUMAN	Q14203 homo sapien
24	78.5	7.1	397	1 MPK4_MOUSE	P47809 m dual spec
25	78.5	7.1	399	1 MPK4_HUMAN	P45985 homo sapien
26	78.5	7.1	791	1 TEX_EORPE	Q45388 bordetella
27	78.5	7.1	1281	1 DYNA_MOUSE	O08788 mus musculu
28	78.5	7.1	1427	1 REST_HUMAN	P30622 homo sapien
29	78	7.0	405	1 FLIK_SALTY	P26416 salmonella
30	78	7.0	1959	1 MYSN_CHICK	P14105 gallus gall
31	78	7.0	3587	1 SRFL_BACSU	P27206 bacillus su
32	77.5	7.0	487	1 AMPL_BOVIN	P00727 bos taurus
33	77.5	7.0	799	1 CAD8_MOUSE	P97291 mus musculu
34	77.5	7.0	962	1 IF3A_MAIZE	O9xhr2 zea mays (m

```

35 77 6.9 164 1 GRP_ZYMMO
36 77 6.9 689 1 ARK1_BOVIN
37 77 6.9 768 1 TRFG_EMENI
38 77 6.9 1947 1 MYSC_CAEEL
39 77 6.9 3567 1 ERY2_SACER
40 76.5 6.9 757 1 YHGF_NEIME
41 76.5 6.9 1204 1 POL_MLVFF
42 76.5 6.9 1301 1 DDX9_CAEEL
43 76 6.8 395 1 METC_BORAV
44 76 6.8 689 1 ARK1_HUMAN
45 76 6.8 1231 1 YKT3_CAEEL

```

```

P74996 zymomonas m
P21146 bos taurus
P06531 emericeella
P12845 caenorhabdi
Q03132 saccharopol
Q51152 neisseria m
P26809 friend muri
Q22307 caenorhabdi
Q07703 bordetella
P25098 homo sapien
P34314 caenorhabdi

```

ALIGNMENTS

```

RESULT 1
Y552_HUMAN
AC O60299; STANDARD; PRT; 673 AA.
DT 15-JUL-1999 (Rel. 38, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DT 15-FEB-2000 (Rel. 39, Last annotation update)
DE HYPOTHETICAL PROTEIN KIAA0552.
GN KIAA0552.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia;
OC Eutheria; Primates; Catarrhini; Hominiidae; Homo.
[1]
RP SEQUENCE FROM N.A.
RC TISSUE=BRAIN;
RX MEDLINE; 98290545.
RA Nagase T., Ishikawa K.-I., Miyajima N., Tanaka A., Kotani H.,
RA Nomura N., Ohara O.;
RT "Prediction of the coding sequences of unidentified human genes. IX.
RT The complete sequences of 100 new cDNA clones from brain which can
RT code for large proteins in vitro.";
RL DNA Res. 5:31-39(1998).
-- SIMILARITY: TO HUMAN KIAA0341.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AB011124; BAA25478.1;
DR Hypothetical protein; Coiled coil.
KW FT DOMAIN 317 496 COILED COIL (POTENTIAL).
FT DOMAIN 571 639 COILED COIL (POTENTIAL).
FT DOMAIN 60 63 POLY-GLY.
FT DOMAIN 122 125 POLY-SER.
FT DOMAIN 282 290 POLY-SER.
FT DOMAIN 298 306 POLY-GLY.
FT DOMAIN 550 553 POLY-ALA.
SQ SEQUENCE 673 AA; 71790 MW; 3E548EC03A01F770 CRC64;

```

Query Match 7.6%; Score 84.5; DB 1; Length 673;
Best Local Similarity 25.2%; Pred. No. 16;
Matches 54; Conservative 34; Mismatches 67; Indels 59; Gaps 12;

```

QY 15 ADSHGRVSFLG-----SOPSPVETAVQAQLKDLDRSTFRKLLVVGALH 60
: : : : : : : : : : : : : : : : : : : : : : : : : :
Db 287 SSSSMGRPHLGSGGGGLPFAACSPSP-----SALIQELEERLWEK--EQEVAAL- 339
QY 61 GKDCRAVEOGLAS-ANLSEER-----LAVLAGHTHTLQQALRLPPASLKPDAFQE- 112
: : : : : : : : : : : : : : : : : : : : : : : : : :
Db 339 -----RRSLEQSEAAVQLSEERKAWERLAELRQCGSKLQQVAR-----RAQRAQQG 388
QY 112 -ELQELGIPQD--LIGDLASLAFSGQRPLLDLSVA---QQQGSLSLPHVSYFRWRV----D 160

```

```
Db 389 LQQLVRLQDDKKQOEAAARLMRRELEEDKVAACQEQADFLPRIETKWEVCQKAGE 448
Qy 161 VAI-----STSAQSRSLQPSVLMQLKLTGSA 187
Db 449 ISLLKQQLKDSQADYSQKLSEIVGLRSQIREGRA 482

RESULT 2
ID RB6K_HUMAN STANDARD; PRT; 890 AA.
AC O95235;
DT 15-FEB-2000 (Rel. 39, Last sequence update)
DE 15-FEB-2000 (Rel. 39, Last sequence update)
DE RABKININ-6 (RAB6-INTERACTING KINESIN-LIKE PROTEIN) (GG10_2).
GN RAB6KIFL.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Primates; Catarrhini; Hominidae; Homo.
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE-ENDOTHELIAL CELLS;
RX MEDLINE: 99252096.
RA Horrevorts A.J.G., Fontijn R.D., van Zonneveld A.J., de Vries C.J.M.,
RA ten Cate J.W., Fannekoek H.;
RT "vascular endothelial genes that are responsive to tumor necrosis
RT factor-alpha in vitro are expressed in atherosclerotic lesions,
RT including inhibitor of apoptosis protein-1, stannin, and two novel
RT genes.";
RL Blood 93:3418-3431(1999).
CC -!- FUNCTION: INTERACTS WITH GUANOSINE TRIPHOSPHATE (GTP)-BOUND FORMS
CC OF RAB6. MAY ACT AS A MOTOR REQUIRED FOR THE RETROGRADE RAB-6
CC REGULATED TRANSPORT OF GOLGI MEMBRANES AND ASSOCIATED VESICLES
CC ALONG MICROTUBULES. HAS A MICROTUBULE PLUS END-DIRECTED MOTILITY
CC (BY SIMILARITY).
CC -!- SUBCELLULAR LOCATION: GOLGI (BY SIMILARITY).
CC -!- SIMILARITY: BELONGS TO THE KINESIN-LIKE PROTEIN FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AF070672; AAC83230.1;
CC DR PROSITE; PS00411; KINESIN_MOTOR_DOMAIN1; FALSE_NEG.
CC DR PROSITE; PS50067; KINESIN_MOTOR_DOMAIN2; 1.
CC KW Motor protein; Microtubules; ATP-binding; Coiled coil; Golgi stack;
CC Protein transport.
CC FT DOMAIN 70 495 MECHANOCHEMICAL (MOTOR).
CC FT NP_BIND 160 167 ATP (POTENTIAL).
CC FT DOMAIN 611 762 COILED COIL (POTENTIAL).
CC FT DOMAIN 763 890 GLOBULAR (POTENTIAL).
CC SEQUENCE 890 AA; 100277 MW; 6620264615496051 CRC64;

Query Match 7.6%; Score 84.5; DB 1; Length 890;
Best Local Similarity 24.5%; Pred. No. 22;
Matches 64; Conservative 32; Mismatches 98; Indels 67; Gaps 13;

Qy 9 PYLHHPADSHSGRVSLGSPSPVETAVA-----QLKDLDRSTFRK--LKLKLVG--- 58
Db 518 PSLHSHFIKHSLOVS-----PSLEKAKADTGLDDDIENEADISWYGRELLQVVEAMKT 572
Qy 58 -----ALHGKD--CREAVEQLGASNLSEERLAVLLAGTHLLQALRLPPAS 103
Db 573 LLLKERQKLOLEMLRDEINEMVEQRCQWQSEHLDTKELLEEMYEKLNILKES 632
Qy 104 LKPDAPQELQELGIPDLGLDGLASLAFSGSRPLDLSVA--QQQSSSLPHVSVFRWRVDA 162
```

```
Db 633 L-TSFYQBEIQE---ROEKIEEALL---QEARQQSVAAHQOQSGSELA-----LRRSQRLA 681
Qy 163 ISTSAQSRSLQPSVLMQLKLTGSAHFRFEVPI-----AKFQELRYSV 204
Db 682 ASASTQQLQOEYKAKLQOQCKAFLNSTTEELHKYQKMLPPPSAKPFTIDVDKLEEGQKNI 741
Qy 205 ALVLKMAEL-----EKKC 218
Db 742 RLLRTELQKLGLGSLQSAERAC 762

RESULT 3
ID GDB1_WHEAT STANDARD; PRT; 304 AA.
AC P04729;
DT 13-AUG-1987 (Rel. 05, Created)
DT 13-AUG-1987 (Rel. 05, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE GAMMA-GLIADIN B-1 PRECURSOR.
OS Triticum aestivum (Wheat).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC euphyllophytes; Spermatophyta; Magnoliophyta; Liliopsida; Poales;
OC Poaceae; Triticum.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE: 85234522.
RA Okita T.W., Cheesbrough V., Reeves C.D.;
RT "Evolution and heterogeneity of the alpha-/beta-type and gamma-type
RT gliadin DNA sequences.";
RL J. Biol. Chem. 260:8203-8213(1985).
CC -!- FUNCTION: GLIADIN IS THE MAJOR SEED STORAGE PROTEIN IN WHEAT.
CC -!- MISCELLANEOUS: THE GAMMA-GLIADINS CAN BE DIVIDED INTO 3 HOMOLOGY
CC CLASSES. SEQUENCE DIVERGENCE BETWEEN THE CLASSES IS DUE TO
CC SINGLE-BASE SUBSTITUTIONS & TO DUPLICATIONS OR DELETIONS WITHIN OR
CC NEAR DIRECT REPEATS.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; M11077; AAA34285.1;
CC DR PRINTS; PR00208; GLIADGLUTEN.
CC KW Seed storage protein; Repeat; Signal; Multigene family.
CC FT SIGNAL 1 23
CC FT CHAIN 24 304 GAMMA-GLIADIN B-1.
CC SEQUENCE 304 AA; 34252 MW; 807EBF447A59D6D5 CRC64;

Query Match 7.6%; Score 84; DB 1; Length 304;
Best Local Similarity 27.3%; Pred. No. 6.4;
Matches 38; Conservative 18; Mismatches 67; Indels 16; Gaps 4;

Qy 99 LPPA---SLKPDAPQELQELGIPDLGLDGLASLAFSGSRPLDLS---VAQQQSSSLPHV 152
Db 38 LPQQSFSQQPPFSQQQQQPL--PQQPSFSQQPPFSQQQPIILSQPPFSQQQPVLPQQ 95
Qy 153 SYFRWRVDVAISTSAQSRSL-----QPSVLMQLKLTGSAHFRFEVPIAKFQELRYSV 204
Db 96 SPFSQQQQQLVLPQQQQQQQLVQQQIPVQPSVLQQLNPKCVFLQOQCSPVAMPQRLARSQ 155
Qy 205 ALVLKMAELEKCKERKIQ 223
Db 156 MWQSSSCHVMQQCCQQLQ 174

RESULT 4
ID TRAL_HUMAN STANDARD; PRT; 661 AA.
AC Q12931; 075235;
```



```
FT DISULFID 492 559 BY SIMILARITY.
FT DISULFID 525 525 INTERCHAIN (BY SIMILARITY).
FT CARBOHYD 64 64 POTENTIAL.
FT CARBOHYD 329 329 POTENTIAL.
FT VARIANT 12 12 V -> G (IN PMS).
FT VARIANT 49 49 /FTID=VAR_007488.
FT VARIANT 70 70 /FTID=VAR_007484.
FT VARIANT 101 101 L -> P (IN PMS).
FT VARIANT 123 123 /FTID=VAR_007485.
FT VARIANT 167 167 G -> V (IN PMS).
FT VARIANT 185 185 /FTID=VAR_007486.
FT VARIANT 194 194 R -> W (IN PMS).
FT VARIANT 325 325 /FTID=VAR_007487.
FT VARIANT 477 477 Y -> C (IN PMS).
FT VARIANT 515 515 /FTID=VAR_007488.
FT VARIANT 560 560 Q -> E.
FT VARIANT 591 591 /FTID=VAR_007489.
FT VARIANT 640 640 R -> C (IN PMS).
FT VARIANT 670 670 /FTID=VAR_007490.
FT VARIANT 700 700 Q -> R.
FT VARIANT 730 730 /FTID=VAR_007491.
FT VARIANT 760 760 V -> A (IN PMS).
FT VARIANT 790 790 /FTID=VAR_007492.
FT VARIANT 820 820 A -> V (IN REF. 2).
FT VARIANT 850 850 3EFC2EB4FECC364C CRC64;
FT VARIANT 880 880 7.4%; Score 82; DB 1; Length 560;
FT VARIANT 910 910 Best Local Similarity 23.5%; Pred. No. 20;
FT VARIANT 940 940 Matches 54; Conservative 17; Mismatches 67; Indels 92; Gaps 9;
FT VARIANT 970 970 29 PSEPTAVAGLL---KDLDRSTFRKLLKLV---GALHGKDCREAVEQVGASANLSEER 81
DB 170 PGEVITVTRAGLPGASLCFSRTRVYLAVDRPAGNWRGSGGLATLQPRGDSRLSTAR 229
QY 82 LAVLLAG-----DSVAQQGSSLPVSVYFRWRVDVAI--STSQAQSRSL 172
DB 230 LQALLFGDHRCTFRMTPTALLLPRLSEPAFLPAHGOLDTVFPFPPRPSAELESPSPADP 289
QY 89 ---THLLQALRLPPA-----SLKPAFQELQELGIPQDLI--GDLASL--AFGSQ 134
DB 290 FLETFLRLVRLVPPARASAPRLADPDAL-----AGPQGLVNLSDPAALERLLDGE 343
QY 135 RPLL-----DSVAQQGSSLPVSVYFRWRVDVAI--STSQAQSRSL 172
DB 344 EPLLRLRLPTAATGDPAPLHLDFTSAPWATLARRVAAELQAAAAELRSL 393
RESULT 8
ID PUR4_DROME STANDARD; PRT: 1354 AA.
AC P35421;
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DT 01-NOV-1995 (Rel. 35, Last annotation update)
DE PHOSPHORIBOSYLFORMYLGLYCINAMIDINE SYNTHASE (EC 6.3.5.3) (FGAM
DE SYNTHASE) (FORMYLGLYCINAMIDE RIBOTIDE AMIDOTRANSFERASE) (FGARAT
DE (ADENOSINE-2) (FGAMS) (FORMYLGLYCINAMIDE RIBOTIDE SYNTHETASE).
GN ADE2.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
[1]
RN SEQUENCE FROM N.A.
RC STRAIN=OREGON-R;
RX MEDLINE: 94095139.
RA Tjong S.Y.K., Nash D.;
RA "The adenosine2 gene of Drosophila melanogaster encodes a
RT formylglycineamide ribotide amidotransferase.";
RL Genome 36:924-934(1993).
CC -|- CATALYTIC ACTIVITY: ATP + 5'-PHOSPHORIBOSYLFORMYLGLYCINAMIDE +
L-GLUTAMINE + H(2)O = ADP + ORTHOPHOSPHATE + 5'-PHOSPHORIBOSYL-
FORMYLGLYCINAMIDINE + L-GLUTAMATE.
-|- PATHWAY: FOURTH STEP IN DE NOVO PURINE BIOSYNTHESIS.
-|- SIMILARITY: TO E.COLI ENZYME (PURL).
This SWISS-PROT entry is copyrighted. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
EMBL; U00683; AAC46488.1;
DR FLYBASE; FBgn0000052; ade2.
DR PFAM; PF00586; ATRS; 1
KW Purine biosynthesis; Ligase.
SQ SEQUENCE 1354 AA; 148195 MW; 30AF726DFF0353A0 CRC64;
Query Match 7.3%; Score 81.5; DB 1; Length 1354;
Best Local Similarity 23.5%; Pred. No. 65;
Matches 61; Conservative 34; Mismatches 84; Indels 81; Gaps 13;
QY 17 SHSGRVSYFLGSDP-----SPQVTA---VAQLLKDLDRSTFRKLLKLV----- 56
DB 725 SHSGIATSTGTQPLKGLDLPAAWARMCMVAEALSNL---VFVKISELADYKCSGNWMAAK 781
QY 56 ---VGALHGKDCREAVEQVGASANLSEERLAVLLAGTHTLLOQALRLPPASLK----- 106
DB 782 LPGEARMEFDACKELQCL-----EELHAIIDGKDSLSMAKVGGETIKSPGTSLVI 833
QY 106 -----PDAFOEQLQELGIPQDLIGDLASLAFSGSRPLL-----DSVAQ---Q 144
DB 834 SYTAPCDV-----RLKVTPLDKGPGA---GSKTSLWINLENSARLGGSALAQAAYQ 883
QY 145 QGSSLPVSVYFRWRVDVAISTQAQSRSLQPSVLMQI--KLTDGS--AHRFEVPIAFQEL 200
DB 884 QGRDNTNLA---RSDVLGKAFVATQSLGDLGLQAGHDVSDGGLVLCVLEMAIGLSGL 939
QY 201 RYSVALVLEMAELEKKCEER 220
DB 940 RVDLSEPLAKLNFKDSVEK 959
RESULT 9
ID ACCA_HAEIN STANDARD; PRT: 315 AA.
AC P43872;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 01-NOV-1995 (Rel. 32, Last annotation update)
DE ACETYL-COENZYME A CARBOXYLASE CARBOXYL TRANSFERASE SUBUNIT ALPHA
DE (EC 6.4.1.2).
GN ACCA OR H10406.
OS Haemophilus influenzae.
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC Haemophilus.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=RD / KW20;
RX MEDLINE: 95350630.
RA Fleischmann R.D., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,
RA McKenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,
RA Scott J.D., Shirley R., Liu L.-I., Glodek A., Kelley J.M.,
RA Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,
RA Utterback T.R., Hanna M.C., Nguyen D.T., Saudek D.M., Brandon R.C.,
RA Fine L.D., Fritchman J.L., Fuhrmann J.L., Geoghegan N.S.M.,
RA Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,
RA Venter J.C.;
RT "Whole-genome random sequencing and assembly of Haemophilus
influenzae Rd.";
```

```

Science 269:496-512(1995).
-!- FUNCTION: THIS PROTEIN IS A COMPONENT OF THE ACETYL COENZYME A
CARBOXYLASE COMPLEX; FIRST, BIOTIN CARBOXYLASE CATALYZES THE
CARBOXYLATION OF THE CARRIER PROTEIN AND THEN THE TRANSCARBOXYLASE
TRANSFERS THE CARBOXYL GROUP TO FORM MALONYL-COA (BY SIMILARITY).
-!- CATALYTIC ACTIVITY: CARBOXYBIOTIN CARBOXYL CARRIER PROTEIN +
ACETYL-COA = BIOTIN CARBOXYL CARRIER PROTEIN + MALONYL-COA.
-!- PATHWAY: FIRST STEP IN LONG-CHAIN FATTY ACID SYNTHESIS.
-!- SUBUNIT: ACETYL-COA CARBOXYLASE IS AN HETEROHEXAMER OF BIOTIN
CARBOXYL CARRIER PROTEIN, BIOTIN CARBOXYLASE AND THE TWO SUBUNITS
OF CARBOXYL TRANSFERASE IN A 2:2 COMPLEX (BY SIMILARITY).
-!- SIMILARITY: TO THE C-TERMINUS OF MAMMALIAN PROPIONYL-COA
CARBOXYLASE BETA CHAIN.
-----
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
-----
EMBL: U32724; AAC22065.1; -
TIGR: H10406; -
PRINTS: PR01069; ACCCTRFRASEA.
KW Fatty acid biosynthesis; Ligase.
SQ SEQUENCE 315 AA; 35126 MW; E3B0B498A2686B23 CRC64;

Query Match 7.2%; Score 80.5; DB 1; Length 315;
Best Local Similarity 25.8%; Pred. NO. 13;
Matches 61; Conservative 28; Mismatches 74; Indels 73; Gaps 12;

QY 12 HHPADSHSGRVSLFSGSPSEVTAVALLLKDLDRSTFRKLLKLVYALHGK-----DC 64
DB 114 HOKGRSVKEVQNRNGMAPE-----GYRAKRLMMEAEFRKLPITITFD 159
QY 65 REAVQLGASANLSEERLAVLLAGHTLLIQALRPPASLKPDAPQEEIQELGIPQ--DL 122
DB 160 PGAYPGICA-----GQAEIARNLR-----EMAQLTVPVICTV 195
QY 123 IGDLAS-----LAFGSRPLDSVAQGGSLPHVS-----YFWRVVDVAISQAQSRSLQ 174
DB 196 IGEQSGGALAIG-----VQDKYNMLQYSTYVISPEGCASILWKAERASTAAEYWGTLA 251
QY 175 SVLMOLKLTLD-----GSAHFEVPIAKFOELYSVALVKE--MAELEKKCKERL 222
DB 252 SRLKEINLDSIVQPLGHAHRSYLEIAENLKR-----LKEDLAELDELSKEEL 301

RESULT 10
POL_MLVRD
ID POL_MLVRD STANDARD; PRT; 1196 AA.
AC P11227;
DT 01-JUL-1989 (Rel. 11, Created)
DT 01-JUL-1989 (Rel. 11, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE POL POLYPROTEIN (CONTAINS: PROTEASE (EC 3.4.23.-); REVERSE
TRANSCRIPTASE (EC 2.7.7.49); RIBONUCLEASE H (EC 3.1.26.4)).
GN POL.
OS Radiation murine leukemia virus.
OC Viruses; Retroviruses; Retroviridae; Mammalian type C retroviruses.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 87207680.
RA Merregert J., Janowski M., Reddy E.P.;
RT "Nucleotide sequence of a radiation leukemia virus genome.";
RL Virology 158:88-102(1987).
CC -!- PTM: SPECIFIC ENZYMATIC CLEAVAGES IN VIVO YIELD NATURE PROTEINS.
CC -!- MISCELLANEOUS: THIS PROTEIN IS SYNTHESIZED AS A GAG-POL
POLYPROTEIN.
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY A2; ALSO KNOWN AS THE
RETROPEPSIN FAMILY.
-----
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
-----
EMBL: K03363; -; NOT_ANNOTATED_CDS.
PIR: B26183; GNMVRV.
DR HSPP; P03355; IMML.
DR PROSITE; PS00141; ASP_PROTEASE; 1.
DR PFAM; PF00075; rnaaseh; 1.
DR PFAM; PF00077; rvp; 1.
DR PFAM; PF00078; rvt; 1.
DR PFAM; PF00552; Integrase; 1.
DR PFAM; PF00665; rve; 1.
KW Hydrolase; Transferase; RNA-directed DNA polymerase;
KW Aspartyl protease; Endonuclease; Polyprotein.
FT CHAIN 1 103 PROTEASE.
ET ACT_SITE 27 27 BY SIMILARITY.
SQ SEQUENCE 1196 AA; 133322 MW; 9549253FCEEB5306 CRC64;

Query Match 7.2%; Score 80.5; DB 1; Length 1196;
Best Local Similarity 24.2%; Pred. No. 67;
Matches 54; Conservative 30; Mismatches 84; Indels 55; Gaps 10;

QY 33 VTAVAQLKDLDRSTFRKLLKLVYALHGKDCREAVEQLGASANLSEERL----AVLAG 88
DB 533 VAAIATVTKDAGKLTMGQ--PLVILAPHAV---EALVQKPPDRWLSNARMTYQAMLLDT 587
QY 89 THTLLQALRLPPASLKPDAPQEEIQELGIPQDLIGDLASLAFSGSRPLDSV----- 142
DB 588 DRVQFGPVVALNPATLP-----LPEGAPHDCLEILAE--THGTEPLDQPPDADHT 640
QY 142 -----AQOQSSSLPHSVFWRVVDVAISQAQSRSLQPSVLMQLKLTGDSAH 189
DB 641 WYDQSSFLQSGQRKAGAAVTEVIWARALPAGTSQAQRAEL-IALTOALKMAEGK--R 697
QY 190 FEVPIAKFOELYSVALVKE-----LVLKMAELEKKCKE 219
DB 698 LNV-----YTDTRYAFATAHGEIYKRRGLLTSEGETKKNKSE 736

RESULT 11
DINA_RAT
ID DINA_RAT STANDARD; PRT; 1280 AA.
AC P28023;
DT 01-AUG-1992 (Rel. 23, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE DYNACTIN, 150 KD ISOFORM (150 KD DYNEIN-ASSOCIATED POLYPEPTIDE)
DE (DP-150) (P150-GLUED).
GN DCTN1.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-SPRAGUE-DAWLEY;
RX MEDLINE; 91260877.
RA Holzbaur E.L.F., Hammarback J.A., Paschal B.M., Kravitt N.G.,
RA Pfister K.K., Vallee R.B.;
RT "Homology of a 150K cytoplasmic dynein-associated polypeptide with
the Drosophila gene Glued.";
RL Nature 351:579-580(1991).
RN [2]
RP REVISIONS.
RA Holzbaur E.L.F., Hammarback J.A., Paschal B.M., Kravitt N.G.,
RA Pfister K.K., Vallee R.B.;
RL Submitted (DEC-1996) to the EMBL/GenBank/DBJ databases.
-----
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
-----
EMBL: K03363; -; NOT_ANNOTATED_CDS.
PIR: B26183; GNMVRV.
DR HSPP; P03355; IMML.
DR PROSITE; PS00141; ASP_PROTEASE; 1.
DR PFAM; PF00075; rnaaseh; 1.
DR PFAM; PF00077; rvp; 1.
DR PFAM; PF00078; rvt; 1.
DR PFAM; PF00552; Integrase; 1.
DR PFAM; PF00665; rve; 1.
KW Hydrolase; Transferase; RNA-directed DNA polymerase;
KW Aspartyl protease; Endonuclease; Polyprotein.
FT CHAIN 1 103 PROTEASE.
ET ACT_SITE 27 27 BY SIMILARITY.
SQ SEQUENCE 1196 AA; 133322 MW; 9549253FCEEB5306 CRC64;
```



```
DE PROTEIN 174).
GN RAB6KIFL.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BALB/C; TISSUE=TESTIS;
RX MEDLINE; 98101856.
RA Echard A., Jollivet F., Martinez O., Lacapere J.-J., Rousselet A.,
RA Janoueix-Lerosey I., Goud B.;
RT "Interaction of a Golgi-associated kinesin-like protein with Rab6.";
RL Science 279:580-585(1998).
CC -!- FUNCTION: INTERACTS WITH GUANOSINE TRIPHOSPHATE (GTP)-BOUND FORMS
CC OF RAB6. MAY ACT AS A MOTOR REQUIRED FOR THE RETROGRADE RAB-6
CC REGULATED TRANSPORT OF GOLGI MEMBRANES AND ASSOCIATED VESICLES
CC ALONG MICROTUBULES. HAS A MICROTUBULE PLUS END-DIRECTED MOTILITY.
CC -!- SUBCELLULAR LOCATION: GOLGI.
CC -!- TISSUE SPECIFICITY: UBIQUITOUSLY EXPRESSED, WITH HIGHEST LEVELS IN
CC SPLEEN AND TESTIS.
CC -!- SIMILARITY: BELONGS TO THE KINESIN-LIKE PROTEIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; Y09632; CAA70845.1; .
DR MGD; MGI:1201682; RAB6KIFL.
DR PRINTS; PR00380; KINESINHEAVY.
DR PROSITE; PS00411; KINESIN_MOTOR_DOMAIN1; FALSE_NEG.
DR PROSITE; PS00067; KINESIN_MOTOR_DOMAIN2; 1.
DR PFAM; PF00225; kinesin; 1.
KW Motor protein; Microtubules; ATP-binding; Coiled coil; Golgi stack;
KW Protein transport.
FT DOMAIN 69 494 MECHANOCHEMICAL (MOTOR).
FT NP_BIND 159 166 ATP (POTENTIAL).
FT DOMAIN 559 804 COILED COIL (POTENTIAL).
FT DOMAIN 805 887 GLOBULAR (POTENTIAL).
FT SEQUENCE 887 AA; 99876 MW; 7143CCB261A4EA15 CRC64;
SQ
```

Query Match 7.2%; Score 80; DB 1; Length 887;
Best Local Similarity 22.7%; Pred. No. 51;
Matches 58; Conservative 35; Mismatches 105; Indels 58; Gaps 10;

```
QY 9 PYLHPADSHSGRV-----SFLGSP-----SPEVTAVAQLKDLDRST 47
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
517 PSLHSFIKKHSPQVPGGLEKEDKADSDLEADVSIVYKKEELQVVEAMKALLKE 576
QY 48 FRKLLKLVGALHGDCREAVEQLGASANLSEERLAVLLAGHTLLOQALRLPPASLKP 107
Db ::||| ::||| ::||| ::||| ::||| ::||| ::||| ::||| ::|||
577 RQELQLKEI-OLRBEICNEMVEQMQQRCQWCSELDNQKELMEELYEKLKILKESL-TT 634
QY 108 AFQELQELGIPQDLIGDASLAFSGRPLDLSVAQQGSSLPVSVFRWVDVAISTSA 167
Db ::||| ::||| ::||| ::||| ::||| ::||| ::||| ::||| ::|||
635 FYQEQIQE---RDEKIELETL---QEAQKQPAQQSGG-----LSLLRSQRLAASAST 684
QY 168 QSRSLQPSVLMQLK-----TDGSAHREFVPI-----AKFOELRYSVALVVK 209
Db |::|| |::|| |::|| |::|| |::|| |::|| |::|| |::||
685 QQFOEVRKAELEQCKTELSSSTAELHKYQQVLKPPPPAKPFTIDVDKLEEGQKNIRLRT 744
QY 210 EMAL-----EKKC 218
Db |::|| |::|| |::|| |::|| |::|| |::|| |::|| |::||
745 ELQKLGSLSQAERAC 760
```

"

"

QY 115 ELGIPQDLIGLASLFGSQRPPLDSVAQQGSSLPVSYFR-----WRVDVAI 163
Db 120 EQTLPOQSSKQSALSPRLSKPEISDSKQQAQALQTPKDSAVRKHSAPSPETQARASISQ 179
QY 164 STASQSRSLQ-----SVLMQLKLDGSAHREVPIAKFOELRYVALVKEMALEK 216
Db 180 ASSSQSLPQESAPERTLLEOKASSFS-----PLSOFSAEKQKEALTTSKSHLYK 233
QY 217 KCERKLQ 223
Db 234 ERDQDQ 240
RESULT 5
ID Q59152 PRELIMINARY; PRT; 584 AA.
AC Q59152;
DT 01-NOV-1996 (TREMBlrel. 01, Created)
DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
DT 01-NOV-1998 (TREMBlrel. 08, Last annotation update)
DE CELD PROTEIN.
GN CELD.
OS Agrobacterium tumefaciens.
OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
OC Rhizobiaceae; Agrobacterium.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 95164506.
RA MATTHYSSE A.G., WHITE S., LIGHTFOOT R.;
RT "Genes required for cellulose synthesis in Agrobacterium
tumefaciens";
RL J. Bacteriol. 177:1069-1075(1995).
DR EMBL; L38609; AAC1431.1; -
SQ SEQUENCE 584 AA; 65787 MW; 24782CA3 CRC32;

Query Match 8.2%; Score 91; DB 2; Length 584;
Best Local Similarity 25.0%; Pred. No. 9.2;
Matches 53; Conservative 27; Mismatches 76; Indels 56; Gaps 10;
QY 22 VSLGSPSPETAVAQLLKDLRSTFRKLLKLVG-----ALHGKDCREAVE 69
Db 263 VLWFGDQ-----LGALANIIDRQK-KSVLFYITGDEWDKTPSGVLVHGCIRRAIE 315
QY 70 Q-----LG-----ASANLSEERLAVLAGT-----HTLQQAALRLPP 101
Db 316 QGFKYDFLRGNEPKYFFGPEEHKLSCTLFTRSGDNLGTLHPRSVRVYEQALKLYK 375
QY 102 ASLKPDFAQELQELGIPQDLIG---DLASLAFGS---QRPLDSVAQQGSSLPVSYFR 155
Db 376 SGKKSEASTAFQGLSAAPHLGAQFGLANLAFDRGEFRKRPSPFASQQATNP---SSF 433
QY 156 RWRVDVAI-STSAQSRSLQPSVLMQLKLDG 186
Db 434 GWRIGETLCATLRSAPESQTVNVRVPPGS 465

RESULT 6
ID Q93317 PRELIMINARY; PRT; 433 AA.
AC Q93317; Q93384;
DT 01-FEB-1997 (TREMBlrel. 02, Created)
DT 01-NOV-1998 (TREMBlrel. 08, Last sequence update)
DT 01-JAN-1999 (TREMBlrel. 09, Last annotation update)
DE C32A3.3 PROTEIN.
GN C32A3.3.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditida;
OC Rhabditia; Rhabditidae; Rhabditidae; Peloderinae; Caenorhabditis.
RN [1]
RP SEQUENCE FROM N.A.
RX BURTON J.;
RA Submitted (NOV-1996) to the EMBL/GenBank/DBJ databases.

RN SEQUENCE FROM N.A.
RP MEDLINE; 94150718.
RX WILSON R., AINSOUGH R., ANDERSON K., BAYNES C., BERKS M.,
RA CRAXTON M., DEAR S., DU Z., DURBIN R., FAVELLO A., FULTON L.,
RA GARDNER A., GREEN P., HAWKINS T., HILLIER L., JIER M., JOHNSTON L.,
RA JONES M., KERSHAW J., KIRSTEN J., LAISTER N., LATREILLE P.,
RA LIGHTNING J., LLOYD C., MCMURRAY A., MORTIMORE B., O'CALLAGHAN M.,
RA PARSONS J., PERCY C., RIFKIN L., ROOPRA A., SAUNDERS D., SHOWNKEEN R.,
RA SMALDON N., SMITH A., SONNHAMMER E., STADEN R., SULSTON J.,
RA THIERRY-MIEG J., THOMAS K., VAUDIN M., VAUGHAN K., WATERSTON R.,
RA WATSON A., WEINSTOCK L., WILKINSON-SPROAT J., WOHLDMAN P.;
RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
elegans";
RL Nature 368:32-38(1994).
RN [3]
RP SEQUENCE FROM N.A.
RX THOMAS K.;
RL Submitted (FEB-1995) to the EMBL/GenBank/DBJ databases.
DR EMBL; Z81449; CAB03766.1; -
DR EMBL; Z48241; CAB03766.1; JOINED.
DR EMBL; Z48241; CAA88286.1; -
DR EMBL; Z81449; CAA88286.1; JOINED.
SQ SEQUENCE 433 AA; 50444 MW; DB1524B0 CRC32;

Query Match 8.0%; Score 89.5; DB 5; Length 433;
Best Local Similarity 23.6%; Pred. No. 8.2;
Matches 57; Conservative 39; Mismatches 95; Indels 51; Gaps 12;
QY 6 AAAPYLHPADSHSGRVSLGSG-----PSPETAVAQLLKDLRSTFRKLLKLVGAL 59
Db 27 SAVTVLHRPRKLSAFVSAMARAASQPSKOITVVD--VYDLAASIGNDFKLDN-- 83
QY 60 HGKDCREAVEQOLGASANLSEERLAVLAGTHTLQQAALRLPPA--SLKPDFAQELQELG 117
Db 83 YGNECVRGIMPKVISA---LETLEAMAAGNDRENEEIMRLSKRAVERLEQEKHQRNOHLK 139
QY 118 IQDQL-----IGDLASLA---FGSQRPPLDSVAQQGSSLPVSYFRWRVDVAIS 164
Db 140 FEELEQVEKTYRKDDLDLQOMVKSIVNENRNLSTTV-----SSLPN-----HADSPVS 188
QY 165 TSAQSRSLQPSVLMQLKLDGSAHREVPI-----AKFQELRYVALVKEMALE 215
Db 189 TSNREADLK--MLELK-EMSSQKDEIKALOKVDVTCQVENLQNSIEKLIRQNEELL 245
QY 216 KK 217
Db 246 RK 247

RESULT 7
ID Q63714 PRELIMINARY; PRT; 3187 AA.
AC Q63714;
DT 01-NOV-1996 (TREMBlrel. 01, Created)
DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
DT 01-MAY-1999 (TREMBlrel. 10, Last annotation update)
DE GIANTIN (GOLGI COMPLEX-ASSOCIATED PROTEIN OF 364 KD) (GCP364).
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
RN [1]
RP SEQUENCE FROM N.A.
RX TISSUE-KIDNEY;
RA TOKI C., MISUMI Y., FUJIWARA T., SOHDA M., NISHIOKA M., IKEHARA Y.;
RT "Identification and characterization of rat 364-kDa Golgi-associated
protein recognized by autoantibodies from a patient with rheumatoid
arthritis";
RL Cell Struct. Funct. 22:565-577(1997).
CC -!- FUNCTION: MAY BE INVOLVED IN THE FORMATION AND/OR MAINTENANCE OF
CC THE CHARACTERISTIC GOLGI STRUCTURE.

QY 136 -----PLDSVAQOQSSPLHSYFRWRVDVAISTSAQSR---SLOPS 175
 DB 241 AOSKPVITFTLDGV-KGEDSSITY-RFQWGNDSYVNIQARAGEFSLIPS 291

RESULT 10
 Q93343 ID Q93343 PRELIMINARY; PRT; 674 AA.
 AC Q93343;
 DT 01-FEB-1997 (TREMBlrel. 02, Created)
 DT 01-JAN-1998 (TREMBlrel. 05, Last sequence update)
 DT 01-NOV-1999 (TREMBlrel. 12, Last annotation update)
 DE C36B1.9 PROTEIN.
 GN C36B1.9
 OS Caenorhabditis elegans.
 OC Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditida;
 OC Rhabditina; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
 RN [1]
 RP SEQUENCE FROM N.A.
 RA LENNARD N.;
 RL Submitted (SEP-1996) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE; 94150718.
 RA WILSON R., AINSOUGH R., ANDERSON K., BAYNES C., BERKS M.,
 RA BONFIELD J., BURTON J., CONNELL M., COPSEY T., COOPER J., COULSON A.,
 RA CRAXTON M., DEAR S., DU Z., DURBIN R., FAVELLO A., FULTON L.,
 RA GARDNER A., GREEN P., HANKINS T., HILLIER L., JIER M., JOHNSTON L.,
 RA JONES M., KERSHAW J., KIRSTEN J., LAISTER N., LATREILLE P.,
 RA LIGHTNING J., LLOYD C., MCMURRAY A., MORTIMORE B., O'CALLAGHAN M.,
 RA PARSONS J., PERCY C., RIKEN L., ROOPRA A., SAUNDERS D., SHOWNKEEN R.,
 RA SHALDON N., SMITH A., SORHAMMER E., STADEN R., SULSTON J.,
 RA THIERRY-MIEG J., THOMAS K., VAUDIN M., VAUGHAN K., WATERSTON R.,
 RA WATSON A., WEINSTOCK L., WILKINSON-SPROAT J., WOHLDMAN P.;
 RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
 RT elegans";
 RL Nature 368:32-38(1994).
 CC -|- SIMILARITY: CONTAINS A C3HC4-CLASS ZINC FINGER.
 DR EMBL; 280215; CAB02274.1; -
 DR PROSITE; PS00518; ZINC_FINGER_C3HC4; 1.
 DR PFAM; PF00097; zf-C3HC4; 1.
 KW Zinc-finger.
 SQ SEQUENCE 674 AA; 77728 MW; 159F7071 CRC32;

Query Match 7.6%; Score 84.5; DB 5; Length 674;
 Best Local Similarity 21.2%; Pred. No. 39;
 Matches 39; Conservative 30; Mismatches 60; Indels 55; Gaps 7;

QY 53 DCREAVEQLGASANLSEERLAVLLAGTHLLQALRLPPASLKPDAFQELQELG--IPQ 120
 DB 111 ECKRINEQVPATVPINPOLMOIL---TTLVLKVLQTPQDQIPKY-ENFERLGTEIP- 166
 QY 121 DLIGDLASLAFSGORPLDSVAQOQSSPLHSYFRWRVDVAISTSAQSRLOPSVLMOL 180
 DB 166 --TADLVNLTGMLYELHKAFFNA-----ILRGKDKTSTIRPSVW--- 205
 QY 181 KLTDGSAHREFVPIAKFQELRYSVALLKEMA-----ELEKKCER 220
 DB 205 -----HRYDSEFKKIDEVERSMRVLHNVIKLOSGETLPRPAPTITPYWOLDREYR 257
 QY 221 KLQD 224
 DB 258 SLQD 261

RESULT 11
 O96275 ID O96275 PRELIMINARY; PRT; 1558 AA.
 AC O96275;
 DT 01-MAY-1999 (TREMBlrel. 10, Created)
 DT 01-MAY-1999 (TREMBlrel. 10, Last sequence update)

DT 01-MAY-1999 (TREMBlrel. 10, Last annotation update)
 DE RESA-H3 ANTIGEN.
 GN PF00915W.
 OS Plasmodium falciparum.
 OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE; 99021743.
 RA KARDNER M.J., TETTELIN H., CARUCCI D.J., CUMMINGS L.M., ARAVIND L.,
 RA KOONIN E.V., SHALLOO S., MASON T., YU K., FUJII C., PEDERSON J.,
 RA SHEN K., JING J., ASTON C., LAI Z., SCHWARTZ D.C., PERTEA M.,
 RA SALZBERG S., ZHOU L., SUTTON G.G., CLAYTON R., WHITE O., SMITH H.O.,
 RA FRASER C.M., ADAMS M.D., VENTER J.C., HOFFMAN S.L.;
 RT "Chromosome 2 sequence of the human malaria parasite Plasmodium
 RT falciparum";
 RT Science 282:1126-1132(1998).
 DR EMBL; AB001424; AAC71972.1; -
 SQ SEQUENCE 1558 AA; 175658 MW; D49C2605 CRC32;

Query Match 7.6%; Score 84.5; DB 5; Length 1558;
 Best Local Similarity 21.6%; Pred. No. 1.2e+02;
 Matches 52; Conservative 45; Mismatches 79; Indels 65; Gaps 10;

QY 27 SOPSPEVTAVQALLKDLDRSFFKLLKLVGALHGKDCREAVEQLGASANLSEERLAVLL 86
 DB 951 SDSKREET---TSIKDKERD-----VSLVVEVQDDMDDESVEKVLKKNMEE----- 997
 QY 87 AGTHTLQOALRLPPASLKPDAFQELQELG-IPQDLIGDLASL-----AFG 132
 DB 997 -----LMKDAVEINDITSK---LIEETQELNEVEADLKDMEKLEKALSDESKKEID 1048
 QY 133 SQRLPLDSVAQOQ---GSSPLHSYFRWRVDVAISTSAQSRLOPSVLMOLK----- 182
 DB 1049 AKDDTLEKVEIEHDITTLDEVKLVKDEKIEKVSQDKLEEDILKEVEKEIREESE 1108
 QY 182 -----LTDGSAHREFVP---IAKF---OELRYSVALLKEMAELKCKERKIQ 223
 DB 1109 ILEDYKELKXTITDILEEKEIEKHDFEAEIEKLEADILKEVSSLEVEEKKLE 1168
 QY 224 D 224
 DB 1169 E 1169

RESULT 12
 P93792 ID P93792 PRELIMINARY; PRT; 303 AA.
 AC P93792;
 DT 01-MAY-1997 (TREMBlrel. 03, Created)
 DT 01-MAY-1997 (TREMBlrel. 03, Last sequence update)
 DT 01-NOV-1999 (TREMBlrel. 12, Last annotation update)
 DE LOW-MOLECULAR-WEIGHT GLUTENIN STORAGE PROTEIN.
 OS Triticum aestivum (Wheat).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC euphyllophytes; Spermatophyta; Magnoliophyta; Liliopsida; Poales;
 OC Poaceae; Triticum.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-CHEYENNE;
 RA ANDERSON O.D., CASSIDY B., DVORAK J.;
 RL Submitted (JAN-1997) to the EMBL/GenBank/DBJ databases.
 DR EMBL; U86028; AAB48477.1; -
 DR MENDEL; 14673; Triae; 2372; 14673.
 DR PRINTS; PR00208; GLIADGLUTEN.
 SQ SEQUENCE 303 AA; 34197 MW; C53AD656 CRC32;

Query Match 7.6%; Score 84; DB 10; Length 303;
 Best Local Similarity 27.38%; Pred. No. 15;
 Matches 38; Conservative 18; Mismatches 67; Indels 16; Gaps 4;

QY 99 LPPA---SLKPDFAQELQELGIPQDLIGDLASLAFSGORPLDS---VAQOQSSPLPH 152

Db 38 LPPQSFQPPFSQQQOPL--PQPFSQQPPFSQQQOPLSQQPPFSQQQOPLVLPQ 95
Qy 153 SYFRNRVDVAISTSAQSRL-----QPSVLMQKLTDSAHFVPIAKFQELRYSV 204
Db 96 SPFSQQOQLVLPQQOQOVLVQOQIPVQPSVYQOQVLPCKVFLQOQOQSPVAMPORLARSQ 155
Qy 205 ALVLKEMAELEKKCKERKQ 223
Db 156 MWQSSCHVMQOCCQQLQ 174
RESULT 13
O43642 PRELIMINARY; PRT; 649 AA.
ID O43642
AC O43642
DT 01-JUN-1998 (TREMBlrel. 06, Created)
DT 01-JUN-1998 (TREMBlrel. 06, Last sequence update)
DT 01-NOV-1999 (TREMBlrel. 12, Last annotation update)
DE HEAT SHOCK PROTEIN 75.
GN HSP75.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrate; Mammalia;
OC Eutheria; Primates; Catarrhini; Hominidae; Homo.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE: 96347537.
RA CHEN C.F., CHEN Y., DAI K., CHEN P.L., RILEY D.J., LEE W.H.;
RT "A new member of the hsp90 family of molecular chaperones interacts
RT with the retinoblastoma protein during mitosis and after heat shock."
RL Mol. Cell. Biol. 16:4691-4699(1996).
RN [2]
RP SEQUENCE FROM N.A.
RA CHEN C.F., CHEN Y., DAI K., CHEN P.L., RILEY D.J., LEE W.H.;
RL Submitted (JAN-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF043254; AAC02679.1; -.
DR HSP; P02829; 1A88.
DR PFAM: PF00183; HSP90; 2.
DR PRINTS: PR00775; HEATSHOCK90.
KW Heat shock.
SQ SEQUENCE 649 AA; 74017 MW; CCAF787C CRC32;
Query Match 7.6%; Score 84; DB 4; Length 649;
Best Local Similarity 23.8%; Pred. No. 40;
Matches 46; Conservative 39; Mismatches 76; Indels 32; Gaps 10;
Qy 49 RKLKLVGALHGKDCREA-VQOLGASANLSEERLAVLLAGTHTLLQQALRLPPASLXPD 107
Db 43 KKLIDIVARSLYSE--KEVFIRELISNASDALEKL-----RHKLVSQOQALPEMELHQL 94
Qy 108 AFQEE----LQELGI---PDLDIGDLASLAFSGSRPLDSVAQOQSSLPVHSY----FR 156
Db 95 TNAEKTITIQDTGIGTQOELVSNLGTARSQKAFDLQALQNAEASSKIIGQFGVGFY 154
Qy 157 WRVDVAISTSAQSRLQPSVLMQKLTDSAHFVPIAKFQELRYSVLVLKEMAELEK 216
Db 155 SAEFVADRVESYSAAGPSGLQYQWLSGSG-VFE--IAEASGVRTGTRKIII-----HLKS 207
Qy 217 KC-----ERKLD 224
Db 208 DCKEFSSEARVD 220
RESULT 14
O07732 PRELIMINARY; PRT; 790 AA.
ID O07732
AC O07732
DT 01-NOV-1996 (TREMBlrel. 01, Created)
DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
DT 01-NOV-1996 (TREMBlrel. 01, Last annotation update)
DE CHROMOSOME IV READING FRAME ORF YDL239C.
OS Saccharomyces cerevisiae (Baker's yeast).

OC Eukaryota; Fungi; Ascomycota; Hemiascomycetes; Saccharomycetales;
OC Saccharomycetaceae; Saccharomycetes.
RN [1]
RP SEQUENCE FROM N.A.
RA ALT-MOERBE J., SCHNEIDER C., MORO M.;
RL Submitted (JUL-1996) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA MIPS;
RL Submitted (JUL-1996) to the EMBL/GenBank/DBJ databases.
DR EMBL: Z74287; CAA98819.1; -.
SQ SEQUENCE 790 AA; 91739 MW; 6974F721 CRC32;
Query Match 7.6%; Score 84; DB 3; Length 790;
Best Local Similarity 21.4%; Pred. No. 52;
Matches 57; Conservative 41; Mismatches 86; Indels 82; Gaps 12;
Qy 11 LHPADSHSRVSLGSPSPSVTAVAL-----LKDLSRSTFRKLLKLVGAL-- 60
Db 517 LHFQOQSHKIKYQKE--QALTKEQOFRLEKRRWHDIILNKEENFQKLKSLKGLIL 573
Qy 60 -----HGKDCREAVEQIGASANL-----SEERLAVLLAGTHTL 92
Db 574 SEKTQNAEDKNDYMEHQEIVEKQNALIASRWSTQIQESENTHKITDELAKQS-- 633
Qy 93 LQQAALRPPA--SLKPDFAFOELQELGIPQDLIGDLASLAFSGSRPLDSVAQOQSSLP 150
Db 633 --ETLKEEITLSLKEDVFOEKLN---LKKLYG-----PSTELNFTVGKSPF 676
Qy 151 HVS---YFRWRVDVAIS--TSAQSRSLQPSVLMQKLTDSAHFVPIAKFQ----- 199
Db 677 HITREKYSGLDILDTLTYVQSNLIKLLIVL-----DIPKRTFLKIVPTIVI 726
Qy 199 ELRYSVALVLKEMAELEKKCKERKLD 224
Db 727 QRCLELLTKFANDLNLKVFQKLD 752
RESULT 15
O96936 PRELIMINARY; PRT; 4151 AA.
ID O96936
AC O96936
DT 01-MAY-1999 (TREMBlrel. 10, Created)
DT 01-MAY-1999 (TREMBlrel. 10, Last sequence update)
DT 01-NOV-1999 (TREMBlrel. 12, Last annotation update)
DE GROOVIN (FRAGMENT).
GN GRV.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE: 99054752.
RA STRUMPF D., VOLK T.;
RT "Kakapo, a novel cytoskeletal-associated protein is essential for the
RT restricted localization of the neurogulin-like factor, veip, at the
RT muscle-tendon junction site."
RL J. Cell Biol. 143:1259-1270(1998).
DR EMBL: Y09430; CAA70581.1; -.
DR HSP; P07751; 1AJ3.
SQ NON_TER 1
SQ SEQUENCE 4151 AA; 472264 MW; 5EBBEC2 CRC32;
Query Match 7.6%; Score 84; DB 5; Length 4151;
Best Local Similarity 19.18%; Pred. No. 4.7e+02;
Matches 56; Conservative 51; Mismatches 108; Indels 78; Gaps 10;
Qy 1 MSALGAAAPYLHHPADSHSRVSLG-----SQPS-----PEVT----- 35
Db 2430 LGILEQALPLSEHFADSHOGLTAWLDDMEQQTSLRSLMPALRPDQITLQODKNERLLQSTIA 2489

QY 35 -----AVALKLDLDRSTFRKLL-----KLVVGAHKGKDCREAVQLGASA 75
Db 2490 EHKPLDKLNTGEGALVADDDGAKINEILDTDNARYAALRLELRERQQALESALQES 2549
QY 76 NLSEERLAVLLAGTHTLLOQALRPPASLKPDAFOBELQELGIPQDLIGDLASLAFGS-Q 134
Db 2550 SQFSDKLEGLMLRALANTVDQVQNLDPISALFPKIREQIEDNALMDDL-DKRODAFSAVQ 2608
QY 135 RPLDLSVAQQQSSSLPHVSYFRMRV-----DVAISTSAQSRSL-----Op----- 175
Db 2609 RAANDVIKAGNKADPAVRDIKAKLEKLNLNWQVQATKRGSSLDLILSVAPFWKQL 2668
QY 175 -SVLMQLKLTGSAHREVEPIAK-----FQELRYSVALYLKEMAELEK 216
Db 2669 NSVMKTLKLEETLSQEPFAAQPDIKKQQAQALQEIHRHEIDOTKPEVEQVRR 2721

Search completed: May 15, 2000, 05:33:25
Job time: 12451 sec

(TM)

Result No.	Query			DB	ID	Description	Pred. No.
	Score	Match	Length				
1	125	7.8	195	11	Q63829	BUP-5'OF BMT-1 PROVIRA	1.42e-03
2	125	7.8	512	10	Q9ZWR5	FAB3.3 PROTEIN.	1.42e-03
3	117	7.3	676	5	Q27512	NEX-2 PROTEIN.	1.69e-02
4	113	7.0	1576	2	Q05204	RECEPTOR-LIKE HISTIDIN	5.64e-02
5	113	7.0	1576	2	O69351	RECEPTOR-LIKE HISTIDIN	5.64e-02
6	107	6.7	360	2	Q33934	ERYCII	3.27e-01
7	107	6.7	361	2	O34225	DTDP-4-KETO-6-DEOXY-HE	3.27e-01
8	107	6.7	451	2	P22863	HYDROPHETAL 49.7 KD P	3.27e-01
9	108	6.7	463	2	Q07346	PMGA, COMPLETE CDS.	4.37e-01
10	107	6.7	463	2	P73374	HYDROPHETAL 51.5 KD P	3.27e-01
11	106	6.6	902	4	O43632	SPINDLE POLE BODY PROT	4.36e-01
12	105	6.5	678	5	O96777	CYCLOC NUCLEOTIDE AND	5.81e-01
13	104	6.5	711	11	Q9WV07	LIPOXYGENASE-3.	7.71e-01
14	103	6.4	327	2	Q9ZBR5	PUTATIVE REGULATORY PR	1.02e+00
15	102	6.4	400	2	O53544	MCE-FAMILY PROTEIN.	1.35e+00
16	103	6.4	764	4	Q9V551	VANILLOID RECEPTOR-LIK	1.02e+00
17	103	6.4	764	4	Q9V670	VANILLOID RECEPTOR-LIK	1.02e+00
18	102	6.4	1116	2	P30707	DNA POLYMERASE III SUB	1.35e+00
19	101	6.3	831	5	O94979	TSH, FSH, LH/CG RECEPT	1.79e+00
20	100	6.2	294	2	O44382	INNER MEMBRANE PROTEIN	2.36e+00

OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC eukaryotes; Spermatophyta; Magnoliophyta; eudicotyledons;
 OC core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae;
 OC Arabidopsis.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CV. COLUMBIA;
 RA FEDERSPIEL N.A., PALM C.J., CONWAY A.B., KURTZ D.B., CONWAY A.R.,
 RA AU M., ARAUJO R., BUEHLER E., DEWAR K., FENG J., KIM C., LI Y.,
 RA OJ O., OSBORNE B.I., SHINN P., SON H., TORIUMI M., VYSOTSKAIA V.S.,
 RA YU G., ECKER J., THEOLOGIS A., DAVIS R.W.;
 RL Submitted (JAN-1999) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AC004146; AAD10659.1; --
 SQ SEQUENCE 512 AA; 58463 MW; 9C771B39 CRC32;
 Query Match 7.8%; Score 125; DB 10; Length 512;
 Best Local Similarity 33.8%; Pred. No. 1.42e-03;
 Matches 22; Conservative 20; Mismatches 18; Indels 5; Gaps 5;
 Db 346 DNRDEVQVCGQGVPSVEQLSSL-TSLNKVINESLRLYPATLLPRMATEDIKLGLDI 404
 QY 62 ECDRDGVQRLGVSANLPE-EQLGALLAGMHTLLQALRL-PPTSLKPD-TFRD-QLQELC 117
 Db 405 IPKGL 409
 QY 118 IPQDL 122
 RESULT 3
 ID Q27512 PRELIMINARY; PRT; 676 AA.
 AC Q27512;
 DT 01-NOV-1996 (Tremblrel. 01, Created)
 DT 01-NOV-1996 (Tremblrel. 01, Last sequence update)
 DT 01-NOV-1999 (Tremblrel. 12, Last annotation update)
 DE NEX-2 PROTEIN.
 GN NEX-2.
 OS Caenorhabditis elegans.
 OC Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditidae;
 OC Rhabditina; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
 RN [1]
 RP SEQUENCE FROM N.A.
 RA BERKS M.;
 RL Submitted (JAN-1994) to the EMBL/GenBank/DBJ databases.
 [2]
 RP SEQUENCE FROM N.A.
 EX MEDLINE; 94150718.
 RA WILSON R., AINSCOUGH R., ANDERSON K., BAYNES C., BERKS M.,
 RA BONFIELD J., BURTON J., CONNELL M., COPSEY T., COOPER J., COULSON A.,
 RA CRAXTON M., DEAR S., DU Z., DURBIN R., FAVELLO A., FULTON L.,
 RA GARDNER A., GREEN P., HAWKINS T., HILLIER L., JIER M., JOHNSTON L.,
 RA JONES M., KERSHAW J., KIRSTEN J., LAISTER N., LATREILLE P.,
 RA LIGHTNING J., LLOYD C., MCMURRAY A., MORTIMORE B., O'CALLAGHAN M.,
 RA PARSONS J., PERCY C., RIFKEN L., ROOPRA A., SAUNDERS D., SHOWNKEEN R.,
 RA SMAILDON N., SMITH A., SONNHAMMER E., STADEN R., SULSTON J.,
 RA THIERRY-MIEG J., THOMAS K., VAUDIN M., VAUGHAN K., WATERSTON R.,
 RA WATSON A., WEINSTOCK L., WILKINSON-SPROAT J., WOHLDMAN P.,
 RA "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
 RT elegans".
 RL Nature 368:32-38(1994).
 CC -1- FUNCTION: CALCIUM/PHOSPHOLIPID-BINDING PROTEIN WHICH PROMOTES
 CC MEMBRANE FUSION AND IS INVOLVED IN EXOCYTOSIS.
 CC -1- SIMILARITY: TO OTHER PROTEINS OF THE ANNEXIN FAMILY.
 DR EMBL: 229443; CAA82571.1; --
 DR HSP; P79134; IAVC.
 DR PROSITE; PS00223; ANNEXIN; 1.
 DR PFAM; PF00191; annexin; 4.
 DR PRINTS; PR00196; ANNEXIN.
 DR Annexin; Calcium/phospholipid-binding; Repeat.
 KW Annexin; Calcium/phospholipid-binding; Repeat.
 SQ SEQUENCE 676 AA; 74969 MW; 00E36C79 CRC32;
 Query Match 7.3%; Score 117; DB 5; Length 676;
 Best Local Similarity 22.5%; Pred. No. 1.69e-02;
 Matches 16; Conservative 26; Mismatches 25; Indels 4; Gaps 4;

Db 132 EKVALERIVEQKLAESAKNEKDIAWNNIVVANSYVAKSLRPVNIOMPTVGGDTNFEED 191
 QY 134 QRPILDSVAQQQ-GAWLPVHADFRRVDVAISTALSARSLQPSVIMQL-KLSDGSAYRFE 191
 Db 192 IDSPAQRQQL 202
 QY 192 VPT-AKF-QEL 200
 RESULT 4
 ID O69354 PRELIMINARY; PRT; 1576 AA.
 AC O69354;
 DT 01-AUG-1998 (Tremblrel. 07, Created)
 DT 01-AUG-1999 (Tremblrel. 11, Last sequence update)
 DT 01-AUG-1999 (Tremblrel. 11, Last annotation update)
 DE RECEPTOR-LIKE HISTIDINE KINASE.
 GN BPHS.
 OS Rhodococcus erythropolis.
 OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
 OC Actinomycetales; Corynebacterineae; Nocardiaceae; Rhodococcus.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-TA421;
 RA ARAI H., KOSONO S., TAGUCHI K., MAEDA M., FUJI F., CHUNG S., SONG E.,
 RA KUDO T.;
 RT "Two sets of the biphenyl degradation genes on a linear plasmid of
 RT Rhodococcus erythropolis TA421.";
 RL Submitted (MAY-1998) to the EMBL/GenBank/DBJ databases.
 [2]
 RN SEQUENCE OF 1-488 FROM N.A.
 RC STRAIN-TA421;
 RA MEDLINE; 97394954.
 RA KOSONO S., MAEDA M., FUJI F., ARAI H., KUDO T.;
 RT "Three of the seven bphC genes of Rhodococcus erythropolis TA421,
 RT isolated from a termitic ecosystem, are located on an indigenous
 RT plasmid associated with biphenyl degradation.";
 RL Appl. Environ. Microbiol. 63:3282-3285(1997).
 DR EMBL: AB014348; BAA34261.1; --
 DR EMBL: D88015; BAA25609.1; --
 KW Receptor; Kinase; Plasmid; Hypothetical protein.
 SQ SEQUENCE 1576 AA; 171035 MW; 636B3907 CRC32;
 Query Match 7.0%; Score 113; DB 2; Length 1576;
 Best Local Similarity 30.3%; Pred. No. 5.64e-02;
 Matches 30; Conservative 18; Mismatches 44; Indels 7; Gaps 7;
 Db 314 MLDGSGVKSR-LLE-AVSEHAQHSKVTVLRAAFDQAPRLGPFSSVFRDLAAH-LR 370
 QY 39 LLGDLDRSTFRKLKFKVYSLQGEDCRDGVRLGVSANLPEELGALLAGMHTLLQALR 98
 Db 371 THPSELQR-V-REELGELVVVGDOVPELAGA-FGGAP 406
 QY 99 LPTSLKPDTRQLQEL-CIPQDLVGLASVVGSGRP 136
 RESULT 5
 ID O05201 PRELIMINARY; PRT; 1576 AA.
 AC O05201;
 DT 01-JUL-1997 (Tremblrel. 04, Created)
 DT 01-JUL-1997 (Tremblrel. 04, Last sequence update)
 DT 01-JUL-1997 (Tremblrel. 04, Last annotation update)
 DE RECEPTOR-LIKE HISTIDINE KINASE BPHS.
 GN BPHS.
 OS Rhodococcus sp.
 OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
 OC Actinomycetales; Corynebacterineae; Nocardiaceae; Rhodococcus.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=M5;
 RA LABBE D., GARNON J., LAU P.C.K.;
 RL J. Bacteriol. 179:0-0(0).

```
DR EMBL; U85412; AAB52543.1; -.
SQ SEQUENCE 1576 AA; 170944 MW; 2E087C0A CRC32;

Query Match 7.0%; Score 113; DB 2; Length 1576;
Best Local Similarity 30.3%; Pred. No. 5.64e-02;
Matches 30; Conservative 18; Mismatches 44; Indels 7; Gaps 7;

Db 314 MLDGSGVGKSR-LLE-AVSEHAQSKVTYVLRRAAFDQAPARPLGPFSSVFRDLAAH-LR 370
QY 39 LGLDLDRSTFRKLLKLVVSSLOGEDCRGQVRLGVSANLPEQLGALLAGMHTLLQALR 98
Db 371 THPSELQR-V-REELGELIVVGDQVPELAGA-FGQAP 406
QY 99 LPSTLSKPDTRDQLQEL-CIPQDLVGDLSVVFSGSRP 136

RESULT 6
ID O33934 PRELIMINARY; PRT; 360 AA.
AC O33934;
DT 01-JAN-1998 (TREMBlrel. 05, Created)
DT 01-JAN-1998 (TREMBlrel. 05, Last sequence update)
DT 01-NOV-1998 (TREMBlrel. 08, Last annotation update)
DE ERYCII.
GN ERYCII.
OS Saccharopolyspora erythraea (Streptomyces erythraeus).
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Pseudonocardiaceae; Pseudonocardiaceae;
OC Saccharopolyspora.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-NRRL 2338;
RX MEDLINE; 98015410.
RA SUMMERS R.G., DONADIO S., STAVER M.J., WENDT-PIENKOWSKI E.,
RA HUTCHINSON C.R., KATZ L.;
RT "Sequencing and mutagenesis of genes from the erythromycin
RT biosynthetic gene cluster of Saccharopolyspora erythraea that are
RT involved in L-mycarose and D-desosamine production.";
RL Microbiology 143:0-0(0).
DR EMBL; U77454; AAB84066.1; -.
DR PFAM; PF00067; P450; 1.
SQ SEQUENCE 360 AA; 38420 MW; D5E834F6 CRC32;

Query Match 6.7%; Score 107; DB 2; Length 360;
Best Local Similarity 25.2%; Pred. No. 3.27e-01;
Matches 33; Conservative 36; Mismatches 54; Indels 8; Gaps 7;

Db 150 MTAVLGAAGVLR--GAANDARVS-LDAQLSPQQLAVTEAAVAALPADPALRAL-FAGAEM 205
QY 1 MSAVGAATPYLHHPGDSHSGRVSLGALPPE-VAAMARLLGLDRLSTFRKLLKLVVSSL 59
Db 206 TANTVVDAVLAVSAEPGLAE-RIADDPAAQRTVAEVLRLHPA-LHLER-RTATAEVRIG 262
QY 60 QGEDCRDGVQRLGVSANLPEQLGALLAGMHTLLQALRLPPTSLKPDTRDQLQELCIP 119
Db 263 EHVIAGEEVV 273
QY 120 QDLVGDLSVY 130

RESULT 7
ID O54225 PRELIMINARY; PRT; 361 AA.
AC O54225;
DT 01-JUN-1998 (TREMBlrel. 06, Created)
DT 01-JUN-1998 (TREMBlrel. 06, Last sequence update)
DT 01-MAY-1999 (TREMBlrel. 10, Last annotation update)
DE DTDLP-4-KETO-6-DEOXY-HEXOSE 3,4-ISOMERASE.
GN ERYCII.
OS Saccharopolyspora erythraea (Streptomyces erythraeus).
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Pseudonocardiaceae; Pseudonocardiaceae;
OC Saccharopolyspora.
RN [1]
RP SEQUENCE FROM N.A.

RC STRAIN-NRRL2338;
RA SALAH-BEY K., DOUMITH M., MICHEL J.M., HAYDOCK S., CORTES J.,
RA LEADLAY P.F., RAYNAL M.C.;
RL Submitted (JUL-1997) to the EMBL/GenBank/DBJ databases.
DR EMBL; Y14332; CAA74711.1; -.
DR PFAM; PF00067; P450; 1.
KW Isomerase.
SQ SEQUENCE 361 AA; 38507 MW; 5FC27F6A CRC32;

Query Match 6.7%; Score 107; DB 2; Length 361;
Best Local Similarity 25.2%; Pred. No. 3.27e-01;
Matches 33; Conservative 36; Mismatches 54; Indels 8; Gaps 7;

Db 151 MTAVLGAAGVLR--GAANDARVS-LDAQLSPQQLAVTEAAVAALPADPALRAL-FAGAEM 206
QY 1 MSAVGAATPYLHHPGDSHSGRVSLGALPPE-VAAMARLLGLDRLSTFRKLLKLVVSSL 59
Db 207 TANTVVDAVLAVSAEPGLAE-RIADDPAAQRTVAEVLRLHPA-LHLER-RTATAEVRIG 263
QY 60 QGEDCRDGVQRLGVSANLPEQLGALLAGMHTLLQALRLPPTSLKPDTRDQLQELCIP 119
Db 264 EHVIAGEEVV 274
QY 120 QDLVGDLSVY 130

RESULT 8
ID P72863 PRELIMINARY; PRT; 451 AA.
AC P72863;
DT 01-FEB-1997 (TREMBlrel. 02, Created)
DT 01-FEB-1997 (TREMBlrel. 02, Last sequence update)
DT 01-JAN-1999 (TREMBlrel. 09, Last annotation update)
DE HYPOTHETICAL 49.7 KD PROTEIN.
OS Synechocystis sp. (strain PCC 6803).
OC Bacteria; Cyanobacteria; Chroococcales; Synechocystis.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-PCC6803;
RA TABATA S.;
RL Submitted (JUN-1996) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-PCC6803;
RA KANEKO T., SATO S., KOTANI H., TANAKA A., ASAMIZU E., NAKAMURA Y.,
RA MIYAJIMA N., HIROSAWA M., SUGIURA M., SASAMOTO S., KIMURA T.,
RA HOSOUCHI T., MATSUNO A., MURAKI A., NAKAZAKI N., NARUO K., OKUMURA S.,
RA SHIMO S., TAKEUCHI C., WADA T., WATANABE A., YAMADA M., YASUDA M.,
RA TABATA S.;
RT "Sequence analysis of the genome of the unicellular cyanobacterium
RT Synechocystis sp. strain PCC6803. II. Sequence determination of the
RT entire genome and assignment of potential protein-coding regions.";
RL DNA Res. 3:109-136(1996).
DR EMBL; D90901; BAAL6879.1; -.
KW Hypothetical protein.
SQ SEQUENCE 451 AA; 49667 MW; EEC3F698 CRC32;

Query Match 6.7%; Score 107; DB 2; Length 451;
Best Local Similarity 38.6%; Pred. No. 3.27e-01;
Matches 22; Conservative 13; Mismatches 18; Indels 4; Gaps 4;

Db 42 ALHFDPPRLPEVEQLTSGT-PEE-LEALALAVETLYQORLN-PCASLEPSTYPPFV 95
QY 58 SLQGEDCR-DGVORLGVSNLPEQLGALLAGMHTLLQALRLPPTSLKPDTRDQL 113

RESULT 9
ID O07346 PRELIMINARY; PRT; 463 AA.
AC O07346;
DT 01-JUL-1997 (TREMBlrel. 04, Created)
DT 01-JUL-1997 (TREMBlrel. 04, Last sequence update)
DT 01-AUG-1998 (TREMBlrel. 07, Last annotation update)
DE PMGA, COMPLETE CDS.
```



```

RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NMRI: TISSUE-SKIN EPIDERMIS: SQUAMOUS CELL CARCINOMA;
RX MEDLINE: 99296830.
RA KINZIG A., HEIDT M., FUERSTENBERGER G., MARKS F., KRIEG P.;
RT "cDNA cloning, genomic structure, and chromosomal localization of a
RL novel murine epidermis-type lipoxigenase.";
DR ENBL: Y14695; CAB46101.1; -.
SQ SEQUENCE 711 AA; 80578 MW; B9BFE292 CRC32;

Query Match 6.5%; Score 104; DB 11; Length 711;
Best Local Similarity 35.0%; Pred. No. 7.71e-01;
Matches 14; Conservative 14; Mismatches 10; Indels 2; Gaps 2;

Db 567 LCTGELVKYLTALIFNCSAOHAFAFNSGQHDGFWMPNAP 606
QY 116 LCIPODLVGLASVVFQ-S-QRPDLDSVAQQGAWLPHVA 153

RESULT 14
ID Q9ZBT5 PRELIMINARY; PRT; 327 AA.
AC Q9ZBT5
DT 01-MAY-1999 (TrEMBLrel. 10, Created)
DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT 01-MAY-1999 (TrEMBLrel. 10, Last annotation update)
DE PUTATIVE REGULATORY PROTEIN.
GN SCIA9.17C.
OS Streptomyces coelicolor.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2);
RA SAUNDERS D.C., HARRIS D.;
RL Submitted (DEC-1998) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2);
RA BENTLEY S.D., PARKHILL J., BARRELL B.G., RAJANDREAM M.A.;
RL Submitted (DEC-1998) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2);
RX MEDLINE: 97000351.
RA KINASHI H., HOPWOOD D.A.;
RA REDENBACH M., KIESER H.M., DENAPATTE D., EICHNER A., CULLUM J.,
RT "A set of ordered cosmid and a detailed genetic and physical map for
RL the 8 Mb Streptomyces coelicolor A3(2) chromosome.";
RL Mol. Microbiol. 21:77-96(1996).
DR ENBL: AL034446; CAA22387.1; -.
SQ SEQUENCE 327 AA; 35032 MW; A2BC81BD CRC32;

Query Match 6.4%; Score 103; DB 2; Length 327;
Best Local Similarity 28.5%; Pred. No. 1.02e+00;
Matches 35; Conservative 33; Mismatches 47; Indels 8; Gaps 7;

Db 9 DTHRAYALVSGRADVPDLARRLALGERDTERALRLRLEQNGLAQSSARPG--RW-VA 65
QY 16 DSHSGRVSEFLGALPPEVAAMRLGLDLDRSTFRLLKFVSSLOGEDC-RDGVQLRGVS 74

Db 66 AP-PGVALGALLTQORHELERA-EIA-AALLAEYFAAAEPAY-HDLVEVVTGAGAVAQ 121
QY 75 ANLPEQLGALLAGMHTLLQQLRLPPTSLKPDTRDQLQELCIPQDLVGLASVVFQSQ 134

Db 122 RFL 124
QY 135 RPL 137

RESULT 15
ID O53544 PRELIMINARY; PRT; 400 AA.
AC O53544;
```

```

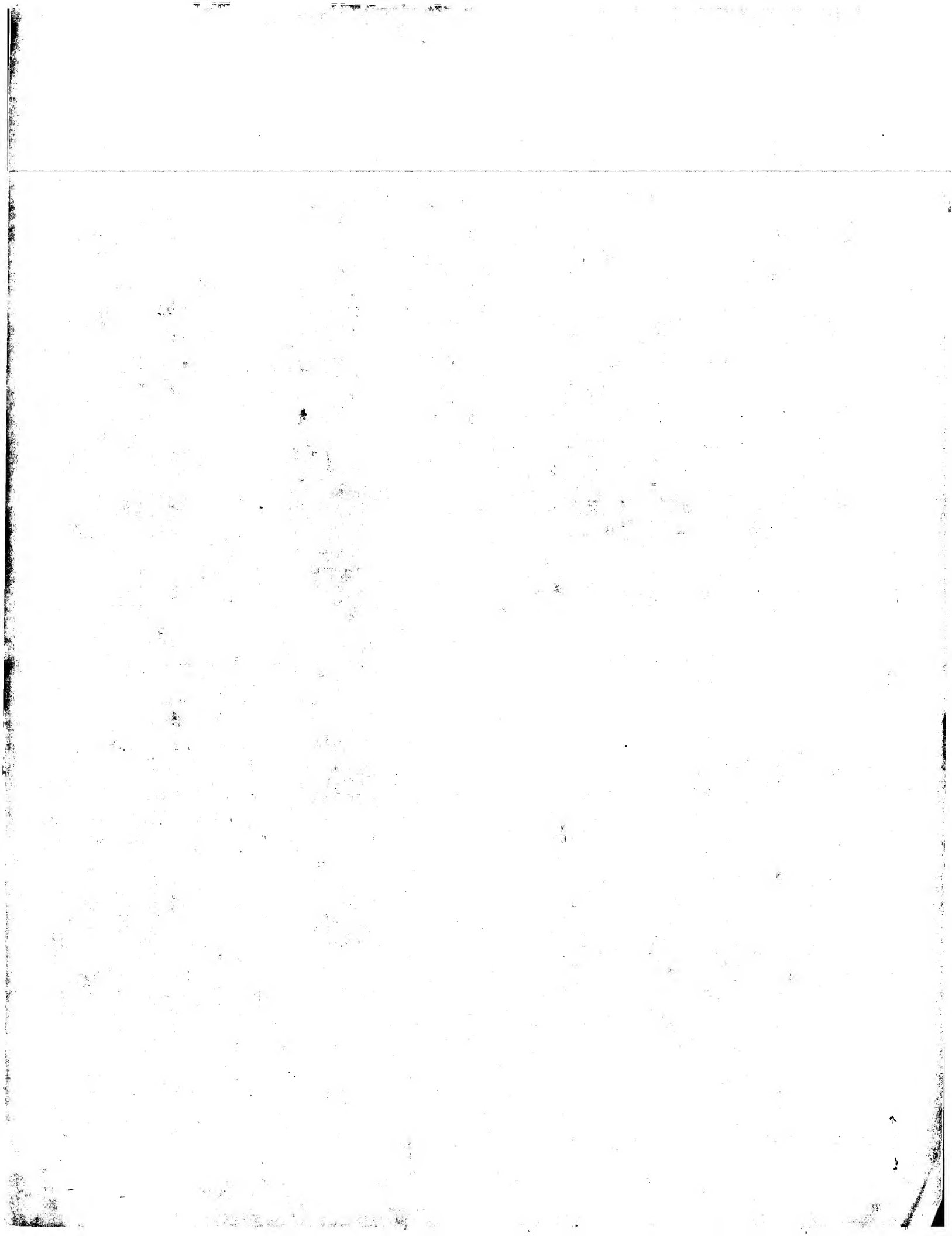
DT 01-JUN-1998 (TrEMBLrel. 06, Created)
DT 01-JUN-1998 (TrEMBLrel. 06, Last sequence update)
DT 01-NOV-1998 (TrEMBLrel. 08, Last annotation update)
DE MCE-FAMILY PROTEIN.
GN MT023.06C.
OS Mycobacterium tuberculosis.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=H37RV;
RA BADCOCK K., CHURCHER C.M.;
RL Submitted (MAR-1998) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=H37RV;
RA COLE S.T., PARKHILL J., BARRELL B.G., RAJANDREAM M.A.;
RL Submitted (MAR-1998) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=H37RV;
RX MEDLINE: 96181548.
RA PHILIPP W.J., POULET S., EIGLMEIER K., PASCOPELLA L.,
RA BALASUBRAMANIAN V., HEYM B., BERGH S., BLOOM B.R., JACOBS W.R. JR.,
RA COLE S.T.;
RT "An integrated map of the genome of the tubercle bacillus,
RT Mycobacterium tuberculosis H37RV, and comparison with Mycobacterium
RT leprae.";
RL Proc. Natl. Acad. Sci. U.S.A. 93:3132-3137(1996).
DR ENBL: AL022022; CAA17736.1; -.
SQ SEQUENCE 400 AA; 42418 MW; 8C21A90C CRC32;

Query Match 6.4%; Score 102; DB 2; Length 400;
Best Local Similarity 26.3%; Pred. No. 1.35e+00;
Matches 31; Conservative 31; Mismatches 48; Indels 8; Gaps 7;

Db 126 LSPNAHVAASQV-QLEVNTLFQSLIDLH-HKIDPLETNATLSALSGLRHGDDLGALLS 183
QY 29 LPPEVAAMARLLGLDLDRST-FRLLKFVSSLOGEDCRDGVQLRGVSANLPEQLGALLA 87

Db 184 GLNTLTRQANPKLP--ALQED-FRKAADVAVNYADAAGDLNTV-FDNLPITNKTIIVDQ 237
QY 88 GMHTLLQQA-LRLPPTSLKPDTRDQLQELCIPQDLVGLASVVFQSQRPDLDSVAQQ 144

Search completed: Fri May 12 10:33:54 2000
Job time : 87 secs.
```



MPERCH_PP protein - protein database search, using Smith-Waterman algorithm

Run on: Fri May 12 10:31:30 2000; MasPar time 34.81 Seconds
Tabular output not generated.
192.185 Million cell updates/sec

Title: >US-09-223-796-4
Description: (1-224) from US09223796.pep
Perfect Score: 1604
Sequence: 1 MSVAGGAATPYLHPGDSHSG.....ALVLKEMADLEKRCERLQD 224

Scoring table: PAM 150
Gap 11

Searched: 82229 seqs, 29864866 residues

Post-processing: Minimum Watch 0%
Listing first 45 summaries

Database: swiss-prot38
1:swisprot

Statistics: Mean 47.835; Variance 94.148; scale 0.508

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
1	110	6.9	2414	1	P300_HUMAN EIA-ASSOCIATED PROTEIN	5.73e-02
2	109	6.8	234	1	NOLW_RHIFR NODULATION PROTEIN NOL	7.74e-02
3	109	6.8	234	1	NOLW_RHISN NODULATION PROTEIN NOL	7.74e-02
4	108	6.7	230	1	HEMX_HABIN PUTATIVE UROPORHYRIN	1.04e-01
5	108	6.7	282	1	LIPH_PSEAE LIPH PROTEIN	1.04e-01
6	108	6.7	302	1	VPJ_BPP2 BASEPLATE ASSEMBLY PRO	1.04e-01
7	106	6.6	340	1	LIMA_PSEAE LIPASE MODULATOR PRECU	1.89e-01
8	103	6.4	2670	1	YAO5_SCHPO PUTATIVE TRANSLATIONAL	4.54e-01
9	102	6.4	3255	1	POLG_LMYE GENOME POLYPROTEIN [CO	6.05e-01
10	101	6.3	81	1	RUXX_METH PUTATIVE SNRNP SM-LIKE	8.05e-01
11	99	6.2	331	1	SIFY_SYNY3 PHENYLALANYL-TRNA SYNT	1.42e+00
12	99	6.2	570	1	FLIF_RHOSH FLAGELLAR M-RING PROTE	1.42e+00
13	100	6.2	784	1	GCF_HUMAN GC-RICH SEQUENCE DNA-B	1.07e+00
14	100	6.2	946	1	YIN7_YEAST HYPOTHETICAL ZINC AMIN	1.07e+00
15	100	6.2	1144	1	R1R1_HSV23 RIBONUCLEOSIDE-DIPHOSP	1.07e+00
16	98	6.1	434	1	YADA_YERPS INVERSAL PRECURSOR [OUT	1.87e+00
17	97	6.0	326	1	GSPK_KLEPN GENERAL SECRETION PATH	2.47e+00
18	97	6.0	661	1	TRAL_HUMAN TUMOR NECROSIS FACTOR	2.47e+00
19	95	5.9	189	1	AROK_SYNY3 SHIKIMATE KINASE (EC 2	4.27e+00
20	94	5.9	306	1	YLIC_ECOLI HYPOTHETICAL ABC TRANS	5.60e+00
21	94	5.9	348	1	HOXV_AZOVI HYDROGENASE EXPRESSION	5.60e+00
22	94	5.9	499	1	CPDF_CANFA CYTOCHROME P450 2D15 (5.60e+00
23	95	5.9	603	1	ALS_RAT INSULIN-LIKE GROWTH FA	4.27e+00

24	94	5.9	1137	1	R1R1_HSV11 RIBONUCLEOSIDE-DIPHOSP	5.60e+00
25	95	5.9	2672	1	GCN1_YEAST TRANSITIONAL ACTIVATO	4.27e+00
26	93	5.8	336	1	YDRI_SCHPO HYPOTHETICAL 38.4 KD P	7.32e+00
27	93	5.8	560	1	YMI9_YEAST HYPOTHETICAL 63.1 KD P	7.32e+00
28	93	5.8	600	1	LAM2_CHICK LAMIN B2	7.32e+00
29	93	5.8	603	1	ALS_MOUSE INSULIN-LIKE GROWTH FA	7.32e+00
30	93	5.8	603	1	Y2N4_CAEEL HYPOTHETICAL ACETYLCHO	7.32e+00
31	93	5.8	836	1	VG26_BPMD2 MINOR TAIL PROTEIN GP2	7.32e+00
32	93	5.8	842	1	PKL2_RAT PROTEIN KINASE C-LIKE	7.32e+00
33	93	5.8	886	1	YFIQ_ECOLI GALACTOSE-INHIBITABLE	7.32e+00
34	93	5.8	1276	1	GILL_ENTHI GENOME POLIPROTEIN [CO	7.32e+00
35	93	5.8	1276	1	POLG_LMYO DFERA PROTEIN	1.24e+01
36	91	5.7	131	1	DFRA_MYXXA INTERLEUKIN-2 PRECURSO	1.24e+01
37	91	5.7	162	1	IL2_CEREL HYPOTHETICAL PROTEIN M	9.54e+00
38	92	5.7	212	1	Y010_MYCPN TRANSCRIPTION FACTOR L	1.24e+01
39	91	5.7	280	1	LBX1_MOUSE HYPOTHETICAL 44.4 KD P	1.24e+01
40	91	5.7	393	1	YJGN_ECOLI FUSION GLYCOPROTEIN PR	9.54e+00
41	92	5.7	538	1	YGLF_MUMPM PUTATIVE VACUOLAR PROT	9.54e+00
42	92	5.7	558	1	YAB3_SCHPO ENV POLYPROTEIN [CONTA	1.24e+01
43	91	5.7	603	1	ENV_RSYP HYPOTHETICAL 84.5 KD P	1.24e+01
44	91	5.7	753	1	YA79_SCHPO HYPOTHETICAL 254.3 KD	9.54e+00
45	92	5.7	2329	1	YS89_CAEEL HYPOTHETICAL 254.3 KD	9.54e+00

ALIGNMENTS

RESULT 1
ID P300_HUMAN STANDARD; PRT: 2414 AA.
AC Q09472;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DE E1A-ASSOCIATED PROTEIN P300.
GN EP300 OR P300.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Primates; Catarrhini; Hominidae; Homo.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE: 95011587.
RA ECKNER R., EWEN M.E., NEWSOME D., GERDES M., DECAPRIO J.A.,
RA LAWRENCE J.B., LIVINGSTON D.M.;
RT "Molecular cloning and functional analysis of the adenovirus E1A-
RT associated 300-kD protein (p300) reveals a protein with properties of
RT a transcriptional adaptor.";
RL Genes Dev. 8:869-884(1994).
CC -!- FUNCTION: PROBABLE TRANSCRIPTIONAL ADAPTOR REQUIRED FOR THE
CC ACTIVITY OF CERTAIN COMPLEX TRANSCRIPTIONAL REGULATORY ELEMENTS.
CC MAY HAVE A FUNCTION IN CELL CYCLE REGULATION. BINDS TO AND MAY BE
CC INVOLVED IN THE TRANSFORMING CAPACITY OF THE ADENOVIRUS E1A
CC PROTEIN.
CC -!- SUBCELLULAR LOCATION: NUCLEAR.
CC -!- SIMILARITY: CONTAINS 1 BROMODOMAIN.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: U01877; AAA18639.1; -
CC MIM: 602700; -
CC PROSITE: PS00633; BROMODOMAIN_1; 1.
CC PROSITE: PS50014; BROMODOMAIN_2; 1.
CC PFAM: PF00439; bromodomain; 1.
CC PFAM: PF00569; Z2; 1.
CC Transcription regulation; Nuclear protein; Bromodomain; Cell cycle.
CC
CC DOMAIN 11 17 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).
CC DOMAIN 797 800 POLY-SER.
CC DOMAIN 1067 1139 BROMODOMAIN.
CC DOMAIN 1519 1526 POLY-GLU.

FT DOMAIN 1572 1818 BINDING REGION FOR EIA ADENOVIRUS.
FT DOMAIN 2066 2069 POLY-GLN.
FT DOMAIN 2190 2195 POLY-GLN.
SQ SEQUENCE 2414 AA; 264143 MW; ED6169CC CRC32;

Query Match 6.9%; Score 110; DB 1; Length 2414;
Best Local Similarity 34.0%; Pred. No. 5.73e-02;
Matches 17; Conservative 13; Mismatches 18; Indels 2; Gaps 2;

Db 2137 OAGVORAGLPQOQOQPPMGHSPQAQO-MNNHNTM-PSQPRDILR 2184
QY 65 RDGVORLGVSNLPEQLGALLAGHMTLLQALRLPPTSLKPDTRDQLQ 114

RESULT 2

ID NOLW_RHIFR STANDARD; PRT; 234 AA.
AC P33212;
DT 01-FEB-1994 (Rel. 28, Created)
DT 01-FEB-1994 (Rel. 28, Last sequence update)
DE 01-OCT-1996 (Rel. 34, Last annotation update)
DE NODULATION PROTEIN NOLW.
GN NOLW.
OS Rhizobium fredii.
OC Plasmid sym.
OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
OC Rhizobiaceae; Sinorhizobium.
[1]
RP SEQUENCE FROM N.A.
RW STRAIN=USDA 257;
RC MEDLINE; 94018604.
RA MEINHARDT L.W., KRISHNAN H.B., BALATTI P.A., PUEPKKE S.G.;
RT Molecular cloning and characterization of a sym plasmid locus that
RT regulates cultivar-specific nodulation of soybean by Rhizobium fredii
RL USDA257.
RL MOL. Microbiol. 9:17-29(1993).
CC -1- FUNCTION: REGULATES CULTIVAR-SPECIFIC NODULATION OF SOYBEAN.
CC -1- SIMILARITY: SOME, TO THE PULD/OUTD/EXED/XPSD FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; L12251; AAB17675.1; -
DR PIR; S35020; S35020.
KW Plasmid; Nodulation; 36 POTENTIAL.
FT TRANSMEM 17
SQ SEQUENCE 234 AA; 25829 MW; 87C798A5 CRC32;

Query Match 6.8%; Score 109; DB 1; Length 234;
Best Local Similarity 27.0%; Pred. No. 7.74e-02;
Matches 34; Conservative 35; Mismatches 49; Indels 8; Gaps 8;

Db 24 LFAGHTTILGATLPSTSYKYVLDQDLSAALQEFGNLKNISVNSAEVKGRIPE 83
QY 85 LLAGHMTLLQALRLPPTSLKPDTRDQL-QELC-IPQDL-VG-DLASVVFSGRPLDS 140
Db 84 LSPREFDLRLDLYLQWYDGVVLYVSA-AKEAQTRMLV-LSSVHFSAFKALDKLDS 141
QY 141 VAQOOG-AWLPHVADFWRVD-VAISALSARSLOPSVLMQLKSDGSAYRFEVPTAKFQ 198
Db 142 DERYPV 147
QY 199 ELRYSV 204

RESULT 3

ID NOLW_RHIN STANDARD; PRT; 234 AA.
AC P35712;
DT 01-NOV-1997 (Rel. 35, Created)

DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE NODULATION PROTEIN NOLW.
GN NOLW OR Y4YD.
OS Rhizobium sp. (strain NGR234).
OG Plasmid sym pNGR234a.
OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
OC Rhizobiaceae; Rhizobium.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 97305956.
RA FREIBERG C.A., FELLAY R., BAIRUCH A., BROUGHTON W.J., ROSENTHAL A.,
RA PERRET X.;
RT "Molecular basis of symbiosis between Rhizobium and legumes.";
RL Nature 387:394-401(1997).
CC -1- SIMILARITY: SOME, TO THE PULD/OUTD/EXED/XPSD FAMILY.
CC -1- SIMILARITY: SOME, TO Y4YD.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AE000107; AAB91943.1; -
KW Nodulation; Plasmid; Transmembrane.
FT TRANSMEM 17
SQ SEQUENCE 234 AA; 25847 MW; 392E7F48 CRC32;

Query Match 6.8%; Score 109; DB 1; Length 234;
Best Local Similarity 27.0%; Pred. No. 7.74e-02;
Matches 34; Conservative 35; Mismatches 49; Indels 8; Gaps 8;

Db 24 LFAGHTTILGATLPSTSYKYVLDQDLSAALQEFGNLKNISVNSAEVKGRIPE 83
QY 85 LLAGHMTLLQALRLPPTSLKPDTRDQL-QELC-IPQDL-VG-DLASVVFSGRPLDS 140
Db 84 LSPREFDLRLDLYLQWYDGVVLYVSA-AKEAQTRMLV-LSSVHFSAFKALDKLDS 141
QY 141 VAQOOG-AWLPHVADFWRVD-VAISALSARSLOPSVLMQLKSDGSAYRFEVPTAKFQ 198
Db 142 DERYPV 147
QY 199 ELRYSV 204

RESULT 4

ID HEMX_HAEIN STANDARD; PRT; 230 AA.
AC P44773;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 01-NOV-1995 (Rel. 32, Last annotation update)
DE PUTATIVE UROPOPHRYN-III C-METHYLTRANSFERASE (EC 2.1.1.107) (UROGEN
DE III METHYLASE).
GN HEMX OR HI0603.
OS Haemophilus influenzae.
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC Haemophilus.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=RD / KW20;
RX MEDLINE; 95350630.
RA KERLAVAGE A.R., BULT C.J., TOMB J.-F., DOUGHERTY B.A., MERRICK J.M.,
RA MCKENNEY K., SUTTON G., FITZHUGH W., FIELDS C.A., GOCAYNE J.D.,
RA SCOTT J.D., SHIRLEY R., LIU L.-I., GLODEK A., KELLEY J.M.,
RA WEIDMAN J.F., PHILLIPS C.A., SPRIGGS T., HEDBLOM E., COTTON M.D.,
RA UTTERBACK T.R., HANNA M.C., NGUYEN D.T., SAUDEK D.M., BRANDON R.C.,
RA FINE L.D., FRITCHMAN J.L., FUHRMANN J.L., GEOGHAGEN N.S.M.,
RA GNEHM C.L., McDONALD L.A., SMALL K.V., FRASER C.M., SMITH H.O.,
RA VENTER J.C.;

DR ENBL; X63391; CAA44998.1; -.

RC STRAIN-TE3285;
RX MEDLINE; 92337414.
RA CHIARA-SIOMI M., YOSHIKAWA K., OSHIMA-HIRAYAMA N., YAMAMOTO K.,
RA SUGABE Y., NAKATANI T., NISHIOKA T., ODA J.;


```
FT CHAIN ? ? NUCLEAR INCLUSION PROTEIN A.
FT CHAIN ? 2977 NUCLEAR INCLUSION PROTEIN B.
FT CHAIN 2978 3255 COAT PROTEIN.
FT NP_BIND 1410 1417 ATP (POTENTIAL).
SQ SEQUENCE 3255 AA; 367621 MW; A5475B9E CRC32;

Query Match 6.4%; Score 102; DB 1; Length 3255;
Best Local Similarity 27.8%; Pred. No. 6.05e-01;
Matches 20; Conservative 14; Mismatches 32; Indels 6; Gaps 6;

Db 1156 WVGVOGVKWLAKSVHYMIPELTNILNVGTLT-LTSLISLV-RFRSLTGQFKMKYKET 1213
QY 148 WLP-HVADFRWRV-DVAISTALSARSLQPSVLMLKLDGSAIRFVPTAKFQELRYSVA 205
Db 1214 LA-RE-EELKR 1223
QY 206 LVLKEMADLEK 217

RESULT 10
ID RUXX.METH STANDARD; PRT; 81 AA.
AC Q26745;
DT 15-JUL-1999 (Rel. 38, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE PUTATIVE SNRNP SM-LIKE PROTEIN.
GN MTH649.
OS Methanobacterium thermoautotrophicum.
OC Archaea; Euryarchaeota; Methanobacteriales; Methanobacteriaceae;
OC Methanobacterium.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-DELTA H.
RX MEDLINE; 98037514.
RA SMITH D.R., DOUCETTE-STAMM L.A., DELOUGHERY C., LEE H.-M., DUBOIS J.,
RA ALDREDGE T., BASHIRZADEH R., BLAKELY D., COOK R., GILBERT K.,
RA HARRISON D., HOANG L., KEAGLE P., LOMW W., FOTHER B., QIU D.,
RA SPADAFORA R., VICARE R., WANG Y., WIERZBOWSKI J., GIBSON R.,
RA JIWANI N., CARUSO A., BUSH D., SAFER H., PAYWELL D., PRABHAKAR S.,
RA MCDUGALL S., SHIMER G., GOVAL A., PIETROVSKI S., CHURCH G.M.,
RA DANIELS C.J., MAO J.-I., RICE P., NOLLING J., REEVE J.N.;
RT "Complete genome sequence of Methanobacterium thermoautotrophicum
RT delah; functional analysis and comparative genomics."
RL J. Bacteriol. 179:1135-1155(1997).
CC -1- SIMILARITY: BELONGS TO THE SNRNP SM PROTEINS FAMILY.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
Db EMBL; A800845; AAB85154.1;
DR PFAM; PF01423; sm; 1.
KW Hypothetical protein.
SQ SEQUENCE 81 AA; 9061 MW; 60D0F649 CRC32;

Query Match 6.38; Score 101; DB 1; Length 81;
Best Local Similarity 34.8%; Pred. No. 8.05e-01;
Matches 23; Conservative 16; Mismatches 19; Indels 8; Gaps 7;

Db 8 RVNVQRPDLALGNSLSPVILKLG-GD-REFR-GVLKS-F-DLH--WNLVNLDAEELEDG 60
QY 158 RVDVAISTALSARSLQPSVLMLKLDGSAIRFVPTAKFQELRYSVALVKEMADLEK- 216
Db 61 EVTRRL 66
QY 217 RCERRL 222

RESULT 11
```

```
ID SYFA_SYNT3 STANDARD; PRT; 331 AA.
AC Q55187;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE PHENYLALANYL-TRNA SYNTHETASE ALPHA CHAIN (EC 6.1.1.20) (PHENYLALANINE-
DE -TRNA LIGASE ALPHA CHAIN) (PHERS).
GN PHER OR SLL0454.
OS Synechocystis sp. (strain PCC 6803).
OC Bacteria; Cyanobacteria; Chroococcales; Synechocystis.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 96127529.
RA KANEKO T., TANAKA A., SATO S., KOTANI H., SAZUKA T., MIYAJIMA N.,
RA SUGIURA M., TABATA S.;
RT "Sequence analysis of the genome of the unicellular cyanobacterium
RT Synechocystis sp. strain PCC6803. I. Sequence features in the 1 Mb
RT region from map positions 64% to 92% of the genome."
RL DNA Res. 2:153-166(1995).
CC -1- CATALYTIC ACTIVITY: ATP + L-PHENYLALANINE + TRNA(PHE) -> AMP +
CC PYROPHOSPHATE + L-PHENYLALANYL-TRNA(PHE).
CC -1- SUBUNIT: Tetramer of two alpha and two beta chains
CC (by similarity).
CC -1- SUBCELLULAR LOCATION: Cytoplasmic.
CC -1- SIMILARITY: BELONGS TO CLASS-II AMINOACYL-TRNA SYNTHETASE FAMILY.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
DR EMBL; D64001; BAA10328.1;
DR HSP; P27001; 1PVS
PROSITE; PS00179; AA-TRNA-LIGASE_II_1; 1.
PROSITE; PS00339; AA-TRNA-LIGASE_II_2; 1.
DR PFAM; PF01409; TRNA-synt_2d; 1.
KW Aminoacyl-TRNA synthetase; Protein biosynthesis; Ligase; ATP-binding.
SQ SEQUENCE 331 AA; 37124 MW; FA9E88E6 CRC32;

Query Match 6.2%; Score 99; DB 1; Length 331;
Best Local Similarity 27.8%; Pred. No. 1.42e+00;
Matches 20; Conservative 23; Mismatches 24; Indels 5; Gaps 4;

Db 33 RVQYLGKGGELSLIKMGKLSAE-ERPKFGAIAINEVKEALQHDLESKANLQNAIEAQ 91
QY 21 RVSFUG--AQLPPEVAMARLLGLDLSRFRKLLKLFVVSLSQGE-DCRDG-VQRLGVSAN 76
Db 92 LAETETIDVTMAG 103
QY 77 LPEEQIGALLAG 88

RESULT 12
ID FLIP_RHOSH STANDARD; PRT; 570 AA.
AC Q53151;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-DEC-1999 (Rel. 39, Last annotation update)
DE FLAGELLAR M-RING PROTEIN.
GN FLIP.
OS Rhodospirillum rubrum (Rhodospirillum rubrum sphaeroides).
OC Bacteria; Proteobacteria; alpha subdivision; Rhodobacter group;
OC Rhodobacter.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-WS8;
RA GOODFELLOW I.G., WOOLLEY K.J., SOCKETT R.E.S.;
RL Submitted (JUL-1996) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: THE M RING MAY BE ACTIVELY INVOLVED IN ENERGY
CC TRANSDUCTION (BY SIMILARITY).
```

CC -1- SUBUNIT: THE BASAL BODY CONSTITUTES A MAJOR PORTION OF THE
CC FLAGELLAR ORGANELLE AND CONSISTS OF FOUR RINGS (L,P,S, AND M)
CC MOUNTED ON A CENTRAL ROD. THE M RING IS INTEGRAL TO THE INNER
CC MEMBRANE OF THE CELL AND MAY BE CONNECTED TO THE FLAGELLAR ROD
CC VIA THE S RING. THE S (SUPRAMEMBRANE RING) LIES JUST DISTAL TO
CC THE M RING. THE L AND P RINGS LIE IN THE OUTER MEMBRANE AND THE
CC PERIPLASMIC SPACE, RESPECTIVELY (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC, MEMBRANE-ASSOCIATED (BY
CC SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE FLIF FAMILY.
CC -----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
CC or send an email to license@lsb-sib.ch).
CC -----
CC EMBL: X98692; CAA67251.1; -
CC DR PFAM: PF01514; YscU_Flif; 1.
CC DR Flagella; Membrane.
CC KW
CC SQ SQUENCE: 570 AA; 60675 MW; 1F898CE2 CRC32;

```

Query Match      6.28;   Score 99;   DB 1;   Length 570;
Best Local Similarity 28.08;   Pred. No. 1.42e+00;
Matches 14;   Conservative 18;   Mismatches 17;   Indels 1;   Gaps 1;

Db 402 PPPLPLPEALKADLERLTSAVGFDADRGDVVTRITAQPFILDTVVPASGW 451
|| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 100 PPTSLLKPFDFRDLQELCIPIQ-DLVGDLASVFGSQRPLLDSSVAOOGRAW 148

```

RESULT	13			
ID	GC F HUMAN	STANDARD;	ERT;	784 AA.
AC	P16383.			
DT	01-AUG-1990 (Rel. 15, Created)			
DT	01-AUG-1990 (Rel. 15, Last annotation update)			
DT	01-JUN-1994 (Rel. 29, Last annotation update)			
DE	GC-RICH SEQUENCE DNA-BINDING FACTOR (GCF) (TRANSCRIPTION FACTOR 9)			
DE	DE (TCF-9).			

TCF9.	
GN	Homo sapiens (Human).
OS	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC	Eutheria; Primates; Catarrhini; Hominoidea; Homo.
[1]	
RN	SEQUENCE FROM N.A.
RP	MEDLINE; 90075226.
RX	KAGEYAMA R., PASTAN I.;
RT	"Molecular cloning and characterization of a human DNA binding factor
RA	that represses transcription.";
RL	Cell 59:815-825(1989).
CC	- - FUNCTION: FACTOR THAT REPRESSES TRANSCRIPTION. IT BINDS TO THE
CC	GC-RICH SEQUENCES (CGGGGG) PRESENT IN THE EPIDERMAL GROWTH FACTOR
CC	RECEPTOR, BETA-ACTIN, AND CALCIUM-DEPENDENT PROTEASE PROMOTERS.
CC	- - SUBCELLULAR LOCATION: NUCLEAR.
CC	- - TISSUE SPECIFICITY: WIDELY EXPRESSED IN HUMAN TISSUES AND CELL

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement. (See <http://www.isb-sib.ch/announce/> or send an email to license@sib-sib.ch).

DR	EMBL; M29204; AAA35598.1; -.
DR	PIR; A33633; A33633.
DR	TRANSFAC; T00320; -.
DR	MIM; 189901; -.
KW	Transcription regulation; Repressor; DNA-binding; Trans-acting factor;
KW	Nuclear protein.

FT	DNA_BIND	1	78
FT	DOMAIN	1	ARG/LYS-RICH (BASIC).
FT	DOMAIN	86	POLY-LYS.
FT	DOMAIN	123	ASP/GLU-RICH (ACIDIC).
FT	DOMAIN	186	LEUCINE-ZIPPER.
FT	DOMAIN	359	LEUCINE-ZIPPER.
FT	DOMAIN	719	LEUCINE-ZIPPER.
FT	DOMAIN	784 AA	FE96504B CRC32;
SQ	SEQUENCE		91001 MW;

Query Match 6.2%; Score 100; DB 1; Length 784;
Best Local Similarity 31.4%; Pred. No. 1.07e+00;
Matches 16; Conservative 12; Mismatches 21; Indels

[illegible]

OC *Saccharomyces cerevisiae* (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Hemiascomycetes; Saccharomycetales;
OC *Saccharomycetaceae*; Saccharomycetes.
OC

[1]
RP SEQUENCE FROM N.A.
RC STRAIN-S28C / AB972;
RA BARRELL B.G., BADCOCK K., BANKIER A.T., BOWMAN S., BROWN D.,
RA CHURCHER C.M., CONNOR R., CORSEY T., DEAR S., DEVLIN K., FRITH
RA GENTLES S., HAMLIN N., HORSNELL T.S., HUNT S., JAGELS K., JOHNS
RA LOUIS E., LYE G., MOULE S., MOULE T., ODELL C., PEARSON D.,
RA RAJANDREAM A.M., RILES L., ROWLEY N., SKELTON J., SMITH V.,
RA WALSH S.V., WHITEHEAD S.;
RL Submitted (DEC-1994) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M1 (ZINC METALLO
CC ALSO KNOWN AS THE PEPN SUBFAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See [http://www.isb-sib.ch/](http://www.isb-sib.ch/announce/)
CC or send an email to license@isb-sib.ch).

CC		ENBL; Z38059; CAA86141.1; -.	
DR	PTR;	S48397; S48397.	
DR	PROSITE;	PS00142; ZINC_PROTEASE; 1.	
DR	PFAM;	PF01433; Peptidase_M1; 1.	
DR	KW	Hypothetical protein; Hydrolase; Metalloprotease; Zinc;	
KW	Aminoacidpeptide.		
FT	METAL	330	ZINC (CATALYTIC) (BY SIMILARITY).
FT	ACT_SITE	331	BY SIMILARITY.
FT	METAL	334	ZINC (CATALYTIC) (BY SIMILARITY).
FT	METAL	353	ZINC (CATALYTIC) (BY SIMILARITY).
SO	SEQUENCE	946 AA; 107722 MW; 568PABF6 CRC32;	

```
Query Match      6.2%;      Score 100;  DB 1;  Length 946;
Best Local Similarity 31.0%;  Pred. No. 1.07e+00;
Matches          22: Conservative 21: Mismatches 20: Indels
```

```

Db      299 GMITIQNHLL-IPPNALANETVREQAQL-IVHELHQWGMGNYISFDS-WESL-WFNES 354
        ||| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
Qv      88 GMHTL-LQOALRPPTSLKPDFTFDIOELCIPDIDVD-IAS-VWFGSORPHLDNSAQO 144

```

DB	355	FATWLACHILE	365
		::: : : :	
QY	145	QGAWLP-HVAD	154

Search completed: Fri May 12 10:32:10 2000
Job time : 40 secs.

```

RESULT 15
ID RTRLHSV23 STANDARD; PRT; 1144 AA.
AC P09853;
DT 01-MAR-1989 (Rel. 10, Created)
DD 01-OCT-1996 (Rel. 34, Last sequence update)
DE 01-OCT-1996 (Rel. 34, Last annotation update)
DS RIBONUCLEOSIDE-DIPHOSPHATE REDUCTASE LARGE CHAIN (EC 1.17.4.1)
DE (RIBONUCLEOTIDE REDUCTASE) (136 KD SUBUNIT).
OS Herpes simplex virus (type 2 / strain 333).
OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC Alphaherpesvirinae; Simplexvirus.
[1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 86144055.
RA SWAN M.A., GALLOWAY D.A.;
RT "Herpes simplex virus specifies two subunits of ribonucleotide
reductase encoded by 3'-coterminal transcripts.";
RL J. Virol. 57:802-808(1986).
[2]
RP SEQUENCE OF 981-1144 FROM N.A.
RX MEDLINE; 84138764.
RA GALLOWAY D.A., SWAN M.A.;
RT "Organization of the left-hand end of the herpes simplex virus type 2
BglII N fragment.";
RL J. Virol. 49:724-730(1984).
[3]
RP SEQUENCE OF 1036-1144 FROM N.A.
RX MEDLINE; 84057718.
RA MC LAUCHLAN J., CLEMENTS J.B.;
RT "DNA sequence homology between two co-linear loci on the HSV genome
which have different transforming abilities.";
RL EMBO J. 2:1953-1961(1983).
CC -!- CATALYTIC ACTIVITY: 2'-DEOXYRIBONUCLEOSIDE DIPHOSPHATE + OXIDIZED
THIOREDOXIN + H(2)O = RIBONUCLEOSIDE DIPHOSPHATE + REDUCED
THIOREDOXIN
CC -!- PATHWAY: FIRST REACTION IN THE DNA REPLICATION PATHWAY.
CC -!- SUBUNIT: HETERODIMER OF A LARGE AND A SMALL CHAIN.
CC -!- SIMILARITY: BELONGS TO THE RIBONUCLEOSIDE DIPHOSPHATE REDUCTASE
LARGE CHAIN FAMILY.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL Outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (see http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
-----
DB EMBL; M12700; AAA45806.1; -.
DR EMBL; X00048; CAA24929.1; -.
DR PIR; A05247; A05247.
DR PROSITE; PS00089; RIBORED_LARGE; 1.
DR PFAM; PF00317; Ribonucleo_red; 1.
KW Oxidoreductase; DNA replication; Early protein.
FT DOMAIN 151 159 POLY-PRO.
FT DOMAIN 224 231 POLY-ASP.
SQ SEQUENCE 1144 AA; 125093 MW; 53E7FDD0 CRC32;
Query Match 6.2%; Score 100; DB 1; Length 1144;
Best Local Similarity 29.7%; Pred.No. 1.07e+00;
Matches 27; Conservative 24; Mismatches 34; Indels 6; Gaps 5;
Db 324 ELTPENAEAVARFLGDVADREP-ALMLEYFORCAREESKRVPPRTFGSPRLTEDDGLL 382
Qy 28 QLPEPVA-AMARLLGD-LDRSTFKLLAFVSSLGQDCRDGVQLGYSANLPPEQLGAL 85
Db 393 NYALAEARRCLLDPPVP--PNATPYHLRE 411
Qy 86 LAGMHTLLQQALRLPPTSLKPDTFRD-QIQE 115

```

MPRLH (TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MPSrch_pp protein - protein database search, using Smith-Waterman algorithm
Run on: Fri May 12 10:34:10 2000; MasPar time 54.28 Seconds
Tabular output not generated. 194.680 Million cell updates/sec

Title: >US-09-223-796-4
Description: (1-224) from US09223796.pep
Perfect Score: 1604
Sequence: 1 MSVGAATPYLHHPGDHSG.....ALVLKEMADLEKRCERRIQD 224

Scoring table: PAM 150
Gap 11

Searched: 142080 segs, 47172406 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: p1r62
1:p1r1 2:p1r2 3:p1r3 4:p1r4

Statistics: Mean 46.697; Variance 102.828; scale 0.454

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
1	117	7.3	676	2	S41022	hypothetical protein
2	110	6.9	2414	2	A54277	transcription adaptor
3	109	6.8	234	2	S35020	nolW protein - Rhizob
4	108	6.7	230	2	D84080	uroporphyrin-III C-met
5	108	6.7	283	2	S25769	lipH protein - Pseudo
6	107	6.7	451	2	S74728	hypothetical protein
7	107	6.7	463	2	S77558	hypothetical protein
8	106	6.6	339	2	S24161	lipase B - Pseudomona
9	102	6.4	400	2	E70805	probable mceA protein
10	102	6.4	1116	2	S77213	DNA-directed DNA poly
11	101	6.3	81	2	C59186	conserved hypothetical
12	99	6.2	331	2	S74410	phenylalanine-tRNA l
13	99	6.2	631	2	T15370	hypothetical protein
14	100	6.2	784	2	A33633	transcription repress
15	100	6.2	946	2	S48397	hypothetical protein
16	98	6.1	368	2	S75923	sensory transduction
17	98	6.1	434	2	S04534	invasin precursor - Y
18	97	6.0	201	2	S11799	pullulanase secretion
19	97	6.0	326	1	S11921	pulk protein - Klebsi
20	97	6.0	479	2	B69764	transcription regulat
21	97	6.0	482	1	A34671	triacylglycerol lipas
22	97	6.0	661	2	A55877	tumor necrosis factor
23	97	6.0	709	2	S75212	comE protein - Synch

24	97	6.0	1111	2	T01239	hypothetical protein	6.85e+00
25	95	5.9	144	2	C70937	hypothetical protein	1.13e+01
26	95	5.9	189	1	S74659	shikimate kinase (EC	1.13e+01
27	95	5.9	221	2	C70936	probable IS1558 trans	1.13e+01
28	94	5.9	306	2	G64820	probable oligopeptidase	1.45e+01
29	95	5.9	333	2	F70678	probable transposase	1.13e+01
30	94	5.9	500	1	JC4157	cytochrome P450 2D, e	1.13e+01
31	95	5.9	520	2	E71416	hypothetical protein	1.13e+01
32	94	5.9	579	2	F70000	two-component sensor	1.45e+01
33	95	5.9	603	2	JC1282	insulin-like growth f	1.13e+01
34	94	5.9	1037	1	WMBEB1	ribonucleoside-diphos	1.45e+01
35	95	5.9	2672	2	A48126	translation activator	1.13e+01
36	93	5.8	467	2	B65020	hypothetical protein	1.86e+01
37	93	5.8	560	2	S50392	hypothetical protein	1.86e+01
38	93	5.8	589	2	S65952	[phosphorylase] phosp	1.86e+01
39	93	5.8	603	2	JC6128	insulin-like growth f	1.86e+01
40	93	5.8	886	2	G65036	hypothetical protein	1.86e+01
41	93	5.8	1292	2	T09229	galactose binding adh	1.86e+01
42	93	5.8	1708	2	F69189	protoporphyrin IX mag	1.86e+01
43	93	5.8	2285	2	T13715	kakapo gene protein i	1.86e+01
44	93	5.8	2396	2	T13714	kakapo gene protein -	1.86e+01
45	93	5.8	4151	2	T13734	groovin gene protein	1.86e+01

ALIGNMENTS

RESULT 1
ENTRY S41022 #type complete
TITLE hypothetical protein T07C4.9 - Caenorhabditis elegans
ORGANISM #formal_name Caenorhabditis elegans
DATE 06-Jan-1995 #sequence_revision 06-Jan-1995 #text_change 09-Sep-1997
ACCESSIONS S41022
REFERENCE S41014
#authors Berks, M.
#submission submitted to the EMBL Data Library, January 1994
#accession S41022
##status preliminary
##molecule_type DNA
##residues 1-676 #label BER
##cross-references EMBL:Z29443; NID:g1067051; PID:g443836
GENETICS
#introns 69/3; 161/1; 208/3; 227/1; 357/1; 505/3
CLASSIFICATION #superfamily annexin repeat homology
FEATURE
373-444 #domain annexin repeat homology #label AX1\
445-516 #domain annexin repeat homology #label AX6\
528-600 #domain annexin repeat homology #label AX3\
604-675 #domain annexin repeat homology #label AX4
SUMMARY #length 676 #molecular-weight 74968 #checksum 2702
Query Match 7.3%; Score 117; DB 2; Length 676;
Best Local Similarity 22.5%; Pred.No. 3.05e-02;
Matches 16; Conservative 26; Mismatches 25; Indels 4; Gaps 4;

Db 132 EKVALERIVEQKLAESAKNEKDIAWNVIVYAKSLRPVINTQMPVGGDTNFEFD 191
QY 134 QREPLDSVAQQQ-GAWLPHFADFRWRVDAISTALARSLOPSVLMQL-KLSDGSAYRFE 191
Db 192 IDSFQAFRQQL 202
QY 192 VPT-AKF-QEL 200

RESULT 2
ENTRY A54277 #type complete
TITLE transcription adaptor protein p300 - human
ORGANISM #formal_name Homo sapiens #common_name man
DATE 09-Sep-1994 #sequence_revision 09-Sep-1994 #text_change 20-Sep-1999
ACCESSIONS A54277
REFERENCE A54277
#authors Eckner, R.; Ewen, M.E.; Newsome, D.; Gerdes, M.; DeCaprio,

[illegible][illegible]


```

Sasamoto, S.; Kimura, T.; Hosouchi, T.; Matsuno, A.;
Muraki, A.; Nakazaki, N.; Naruo, K.; Okumura, S.; Shimpo,
S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.;
Yasuda, M.; Tabata, S.
DNA res. (1996) 3:109-136
#journal
#title Sequence analysis of the genome of the unicellular
cyanobacterium Synecocystis sp. PCC6803. II. Sequence
determination of the entire genome and assignment of
potential protein-coding regions.
#cross-references MID:97061201
#accession S77558
##status nucleic acid sequence not shown; translation not shown
##molecule_type DNA
##residues 1-463 #label KAN
##cross-references EMBL:D90905; GB:AB001339; NID:g1652360; PID:d1018138
PID:g1652493
##note the nucleotide sequence was submitted to the ENBL Data
Library, June 1996
CLASSIFICATION #superfamily hypothetical protein H10333
SUMMARY #length 463 #molecular-weight 51523 #checksum 8580
Query Match 6.7%; Score 107; DB 2; Length 463;
Best Local Similarity 22.3%; Pred. No. 4.96e-01;
Matches 31; Conservative 41; Mismatches 57; Indels 10; Gaps 10;
Db 116 RIGGFTTLPLELLQSPASL-GYRNKATYP-LSRSTKGQVAGYYRKGSHRLVNINQCVP 173
:: | | | | | :: | : : : | : : | | : :
Qy 21 RVSLFLGAQLPPEVAAMARLLGLDRSTFRKLKFVVSSIQEGDCRDGVQL-GVSA-NL 77
:: | | | | | :: | : : : | : : | | : :
Db 174 QDRLNLLTEVKDDIENGWSYDEKKQGKLR-HL-SLRIGQR-TGEMLTLLISAHGK 230
:: | | : : : : : | : | | | | : : : : :
Qy 78 PEELGALLAGMTLLQ-QALRPPTSLKPDPTRDLQELCPDILVGDLASVFGSQRP 136
| | : : | | | : : | : | | | : :
Db 231 LPD-LEEQAQEWLERYPDL 248
| | : : | | | : : | :
Qy 137 LLDSVAQQQGAWLPHVADF 155
| | : : | | | : : | :
RESULT 8
ENTRY S24161 #type complete
TITLE lipase B - Pseudomonas aeruginosa
ORGANISM #formal_name Pseudomonas aeruginosa
DATE 19-Mar-1997 #sequence_revision 19-Mar-1997 #text_change
20-Mar-1998
ACCESSIONS S24161
REFERENCE S24160
#authors Chihara-Stomi, M.; Yoshikawa, K.; Oshima-Hirayama, N.;
Yamamoto, K.; Sogabe, Y.; Nakatani, T.; Nishioka, T.; Oda,
J.
#journal Arch. Biochem. Biophys. (1992) 296:505-513
#title Purification, molecular cloning, and expression of lipase
from Pseudomonas aeruginosa.
#cross-references MID:92337414
#accession S24161
##status preliminary
##molecule_type DNA
##residues 1-339 #label CHI
##cross-references GB:D10048; NID:g216895; PID:d1001403; PID:g216897
SUMMARY #length 339 #molecular-weight 37532 #checksum 1707
Query Match 6.6%; Score 106; DB 2; Length 339;
Best Local Similarity 30.5%; Pred. No. 6.50e-01;
Matches 18; Conservative 21; Mismatches 17; Indels 3; Gaps 2;
Db 207 LSTEKAAAIDL--RASLPEDQGESVLPOQSELOQTAAALQAAGAPEAIRMQEOL 263
| | : : | | | : : | : : | : : | : : | : :
Qy 59 LGGEDCDRGVQRIGVSANLPEQGLLAGMHT-LIQQALRLPPTSLKPDTFRDLQDEL 116
| | : : | | | : : | : : | : : | : : | : :
RESULT 9
ENTRY E70805 #type complete
TITLE probable mce4 protein - Mycobacterium tuberculosis (strain
H37RV)

```


Mon May 15 15:24:26 2000

US-09-223-796-4.rpr

Page 6

. QY ::||: | : :
 145 QGAWLP-HVAD 154

Search completed: Fri May 12 10:35:10 2000
Job time : 60 secs.

MPERCH_PP protein - protein database search, using Smith-Waterman algorithm
Run on: Fri May 12 10:36:18 2000; MasPar time 17.97 Seconds
Tabular output not generated.
161.524 Million cell updates/sec

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

Title: >US-09-223-796-4
Description: (1-224) from US09223796.pep
Perfect Score: 1604
Sequence: 1 MSVAGTAATPYLHPGDSHG.....ALVLKEMADLEKRCERRLQD 224

Scoring table: PAM 150
Gap 11

Searched: 131253 seqs, 12956647 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: a-issued
1:5A_COMB 2:5B_COMB 3:PCT_COMB 4:backfiles1

Statistics: Mean 30.751; Variance 151.832; scale 0.203

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query Match	Length	DB ID	Description	Pred. No.
1	125	7.8	195	2 US-08-822- Sequence 3, Applicatio	2.01e+01
2	122	7.6	195	2 US-08-822- Sequence 1, Applicatio	3.27e+01
3	110	6.9	2414	3 PCT-US95-0 Sequence 2, Applicatio	2.22e+00
4	110	6.9	2414	3 US-08-227- Sequence 2, Applicatio	2.22e+00
5	107	6.7	395	2 US-08-795- Sequence 18, Applicati	3.56e+00
6	107	6.7	395	1 US-08-318- Sequence 18, Applicati	3.56e+00
7	107	6.7	1114	2 US-08-576- Sequence 31, Applicati	3.56e+00
8	100	6.2	448	2 US-09-015- Sequence 1, Applicatio	1.05e+01
9	98	6.1	1064	1 US-08-357- Sequence 2, Applicatio	1.43e+01
10	98	6.1	1064	1 PCT-US95-1 Sequence 2, Applicatio	1.43e+01
11	98	6.1	1064	1 US-09-003- Sequence 5, Applicatio	1.43e+01
12	98	6.1	1082	1 US-08-357- Sequence 10, Applicati	1.43e+01
13	98	6.1	1082	1 PCT-US95-1 Sequence 5, Applicatio	1.43e+01
14	98	6.1	1082	1 US-08-357- Sequence 10, Applicati	1.43e+01
15	98	6.1	1082	1 PCT-US95-1 Sequence 5, Applicatio	1.43e+01
16	98	6.1	1082	1 US-08-357- Sequence 10, Applicati	1.43e+01
17	98	6.1	1082	2 US-09-003- Sequence 5, Applicatio	1.43e+01
18	94	5.9	382	1 US-08-795- Sequence 17, Applicati	2.61e+01
19	94	5.9	382	1 US-08-318- Sequence 17, Applicati	2.61e+01
20	95	5.9	603	1 US-08-190- Sequence 50, Applicati	2.25e+01
21	93	5.8	155	2 US-08-576- Sequence 55, Applicati	3.04e+01
22	92	5.7	335	2 US-08-875- Sequence 3, Applicatio	3.53e+01
23	92	5.7	350	3 PCT-US95-0 Sequence 14, Applicati	3.53e+01

24	92	5.7	350	2 US-08-765- Sequence 14, Applicati	3.53e+01
25	91	5.7	476	1 US-08-565- Sequence 6, Applicatio	4.09e+01
26	91	5.7	490	2 US-08-946- Sequence 2, Applicatio	4.09e+01
27	91	5.7	490	1 US-08-361- Sequence 2, Applicatio	4.09e+01
28	91	5.7	490	1 US-08-565- Sequence 2, Applicatio	4.09e+01
29	91	5.7	3031	1 US-07-689- Sequence 2, Applicatio	4.09e+01
30	88	5.5	454	2 US-09-014- Sequence 9, Applicatio	6.38e+01
31	89	5.5	494	1 US-08-447- Sequence 2, Applicatio	5.51e+01
32	89	5.5	494	1 US-08-453- Sequence 2, Applicatio	5.51e+01
33	89	5.5	494	1 US-08-447- Sequence 2, Applicatio	5.51e+01
34	89	5.5	494	1 US-08-454- Sequence 2, Applicatio	5.51e+01
35	87	5.4	419	1 US-08-385- Sequence 2, Applicatio	7.39e+01
36	86	5.4	560	1 US-07-683- Sequence 1, Applicatio	8.55e+01
37	86	5.4	963	1 US-08-537- Sequence 3, Applicatio	8.55e+01
38	86	5.4	1143	2 US-08-310- Sequence 108, Applicat	8.55e+01
39	86	5.4	1143	3 PCT-US95-0 Sequence 108, Applicat	8.55e+01
40	86	5.4	1144	1 US-08-261- Sequence 2, Applicatio	8.55e+01
41	86	5.4	1144	3 PCT-US95-0 Sequence 4, Applicatio	8.55e+01
42	86	5.4	1144	1 US-08-261- Sequence 4, Applicatio	8.55e+01
43	86	5.4	1144	3 PCT-US95-0 Sequence 2, Applicatio	8.55e+01
44	84	5.2	394	1 US-08-002- Sequence 9, Applicatio	1.14e+02
45	84	5.2	394	1 US-08-002- Sequence 6, Applicatio	1.14e+02

ALIGNMENTS

RESULT 1
ID US-08-822-260-3 STANDARD; PRT; 195 AA.
XX
AC xxxxxx
XX
DT
XX
DE Sequence 3, Application US/08822260
XX
CC Sequence 3, Application US/08822260
CC Patent No. 5830660
CC GENERAL INFORMATION:
CC APPLICANT: Hillman, Jennifer L.
CC APPLICANT: Goli, Surya K.
CC TITLE OF INVENTION: NOVEL TUMORIGENESIS PROTEIN
CC NUMBER OF SEQUENCES: 3
CC CORRESPONDENCE ADDRESS:
CC ADDRESS: Incyte Pharmaceuticals, Inc.
CC STREET: 3174 Porter Drive
CC CITY: Palo Alto
CC STATE: CA
CC COUNTRY: USA
CC ZIP: 94304

COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/822.260
FILING DATE: Herewith
CLASSIFICATION: 514
PRIOR APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Billings, Lucy J.
REGISTRATION NUMBER: 36,749
REFERENCE/DOCKET NUMBER: PF-0247 US
TELEPHONE: 415-855-0555
TELEFAX: 415-845-4166
TELEX:
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 195 amino acids
TYPE: amino acid

CC STRANDEDNESS: single
CC TOPOLOGY: linear
CC IMMEDIATE SOURCE:
CC LIBRARY: GenBank
CC CLONE: 265569
SQ SEQUENCE 195 AA; 22037 MW; 190094 CN;

Query Match 7.8%; Score 125; DB 2; Length 195;
Best Local Similarity 22.2%; Pred. No. 2.01e-01;
Matches 24; Conservative 37; Mismatches 40; Indels 7; Gaps 7;

Db 87 LED-CKDRERI-ELFCTEQNNKNSLETLGSGRSLPHITDVSRLVQIKTNQHKM 144
QY 113 LQELC-IPQDLVGLASVVGSGRPLDLSVAQOQGWLPVADFRWRVDVAISTALARS 171

Db 145 YRPGYLVTLNVDNSOSYPIFNESCNEQLQDLVG-KLKDASKSLER 191
QY 172 LQPSVLQKLSDGSAYRF-EVP-TAKFQELRYVALVKEMA-DLEK 216

RESULT 2
ID US-08-822-260-1 STANDARD; PRT; 195 AA.
XX
AC xxxxxx
XX
DT

Sequence 1, Application US/08822260

Sequence 1, Application US/08822260

Patent No. 5830660

GENERAL INFORMATION:

APPLICANT: Hillman, Jennifer L.

APPLICANT: Goli, Surya K.

TITLE OF INVENTION: NOVEL TUMORIGENESIS PROTEIN

NUMBER OF SEQUENCES: 3

CORRESPONDENCE ADDRESS:

ADDRESSEE: Incyte Pharmaceuticals, Inc.

STREET: 3174 Porter Drive

CITY: Palo Alto

STATE: CA

COUNTRY: USA

ZIP: 94304

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette

COMPUTER: IBM Compatible

OPERATING SYSTEM: DOS

SOFTWARE: FastSeq for Windows Version 2.0

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/822,260

FILING DATE: Herewith

CLASSIFICATION: 514

PRIOR APPLICATION DATA:

APPLICATION NUMBER:

FILING DATE:

ATTORNEY/AGENT INFORMATION:

NAME: Billings, Lucy J.

REGISTRATION NUMBER: 36,749

REFERENCE/DOCKET NUMBER: PF-0247 US

TELECOMMUNICATION INFORMATION:

TELEPHONE: 415-855-0555

TELEFAX: 415-845-4166

TELEX:

INFORMATION FOR SEQ ID NO: 1:

SEQUENCE CHARACTERISTICS:

LENGTH: 195 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

IMMEDIATE SOURCE:

LIBRARY: UTRSN02

CLONE: 2267574

SEQUENCE 195 AA; 22151 MW; 196728 CN;

Query Match 7.6%; Score 122; DB 2; Length 195;
Best Local Similarity 24.8%; Pred. No. 3.27e-01;
Matches 27; Conservative 35; Mismatches 38; Indels 9; Gaps 8;

Db 87 LED-CKDRERI-ELFCTEQNNKNSLEILLGSGRSLPHITDVSRLVQIKTNQHKM 144
QY 113 LQELC-IPQDLVGLASVVGSGRPLDLSVAQOQGWLPVADFRWRVDVAISTALARS 171

Db 145 YRPAVLVTLVQNTDPSVP-EISFSCSMEQLQDLVG-KLKDASKSLER 191
QY 172 LQPSVLQKLK--SDGSAYRFEVP-TAKFQELRYVALVKEMA-DLEK 216

RESULT 3
ID PCT-US95-04682-2 STANDARD; PRT; 2414 AA.
XX
AC xxxxxx
XX
DT

Sequence 2, Application PC/TUS9504682

Sequence 2, Application PC/TUS9504682

GENERAL INFORMATION:

TITLE OF INVENTION: NUCLEIC ACID ENCODING TRANSCRIPTION

NUMBER OF SEQUENCES: 13

CORRESPONDENCE ADDRESS:

ADDRESSEE: Weingarten, Schurgin, Gagnebin & Hayes

STREET: Ten Post Office Square

CITY: Boston

STATE: MA

COUNTRY: US

ZIP: 02109

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: PCT/US95/04682

FILING DATE:

CLASSIFICATION:

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/227,536

FILING DATE: 14-April-1994

CLASSIFICATION:

ATTORNEY/AGENT INFORMATION:

NAME: Holliday C. Heine, Ph.D.

REGISTRATION NUMBER: 34,346

REFERENCE/DOCKET NUMBER: DFCI-308Xq999

TELECOMMUNICATION INFORMATION:

TELEPHONE: (617) 542-2230

TELEFAX: (617) 451-0313

INFORMATION FOR SEQ ID NO: 2:

SEQUENCE CHARACTERISTICS:

LENGTH: 2414 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

SEQUENCE 2414 AA; 264143 MW; 29411911 CN;

Query Match 6.9%; Score 110; DB 3; Length 2414;
Best Local Similarity 34.0%; Pred. No. 2.22e-00;
Matches 17; Conservative 13; Mismatches 18; Indels 2; Gaps 2;

Db 2137 QAGVQRAGLPQQOQLOPPHGMSPQAO-MNNMNTM-PSQFRIILR 2184
QY 65 RDGVQRGLVSANLPEQQLGALLAGMHTLLQALRLPPTSLKPDTRDQLQ 114

	TITLE OF INVENTION:	COMPLEMENTARY DNA ENCODING THE SAME
	NUMBER OF SEQUENCES:	21
	CORRESPONDENCE ADDRESS:	
	ADDRESSEE:	Sughrue, Mion, Zinn, Macpeak & Seas
	STREET:	2100 Pennsylvania Avenue, NW Suite 800
	CITY:	Washington
	STATE:	DC
	COUNTRY:	USA
	ZIP:	20037
	COMPUTER READABLE FORM:	
	MEDIUM TYPE:	Floppy disk
	COMPUTER:	IBM PC compatible
	OPERATING SYSTEM:	PC-DOS/MS-DOS
	SOFTWARE:	PatentIn Release #1.0, Version #1.25
	CURRENT APPLICATION DATA:	
	APPLICATION NUMBER:	US/08/795,303
	FILING DATE:	04-FEB-1997
	CLASSIFICATION:	435
	PRIOR APPLICATION DATA:	
	APPLICATION NUMBER:	08/318,947
	FILING DATE:	06-OCT-1994
	APPLICATION NUMBER:	08/133,530
	FILING DATE:	07-OCT-1993
	ATTORNEY/AGENT INFORMATION:	
	NAME:	Mack, Susan J.
	REGISTRATION NUMBER:	30,951
	REFERENCE/DOCKET NUMBER:	A6462
	TELECOMMUNICATION INFORMATION:	
	TELEPHONE:	(202)293-7060
	TELEFAX:	(202)293-2920
	TELEX:	6491103
	INFORMATION FOR SEQ ID NO:	18:
	SEQUENCE CHARACTERISTICS:	
	LENGTH:	395 amino acids
	TYPE:	amino acid
	TOPOLOGY:	linear
	MOLECULE TYPE:	protein
	SEQUENCE	395 AA; 42522 MW; 753192 CN;
	Query Match	6.7%; Score 107; DB 2; Length 395;
	Best Local Similarity	29.7%; Pred. No. 3.56e+00;
	Matches	27; Conservative 25; Mismatches 33; Indels 6
D _b	180 ELTPENAEAVARFLGDAVDREP-	ALMLEYFCRCARESKRVPPTFGSAPRLTED
	: : : : : : :	:
Q _y	28 QLPEVA-AMARLIGD-LDRSTRKLLKFVSSLOGEDCRDGVRQLGVSANLPES	
	::: : : : : :	:
D _b	239 NTAAEMRRCLGLDPYP--PNAYTPLYHLRE	267
	::: : : : : :	:
Q _y	86 LAGMHTLQQALRPPTSLSKPDTFRD-QIQE	115
RESULT	6	
ID	US-08-318-947A-18	STANDARD; PRT; 395 AA.
XX	xxxxxx	
AC		
XX		
XX		
DT		
XX		
DE	Sequence 18, Application US/08318947A	
CC	Sequence 18, Application US/08318947A	
CC	Patent No. 5798245	
CC	GENERAL INFORMATION:	
CC	APPLICANT:	Anderson, Paul J.
CC	APPLICANT:	Tian, Qingsheng
CC	TITLE OF INVENTION:	TIA-1 BINDING PROTEINS AND ISOLATED
CC	TITLE OF INVENTION:	COMPLEMENTARY DNA ENCODING THE SAME
CC	NUMBER OF SEQUENCES:	21
CC	CORRESPONDENCE ADDRESS:	
CC	ADDRESSEE:	Sughrue, Mion, Zinn, Macpeak & Seas
CC	STREET:	2100 Pennsylvania Avenue, NW Suite 800
CC	CITY:	Washington

```
CC STATE: DC
CC COUNTRY: USA
CC ZIP: 20037
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Floppy disk
CC COMPUTER: IBM PC compatible
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: PatentIn Release #1.0, Version #1.25
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/08/318,947A
CC FILING DATE: 06-OCT-1994
CC CLASSIFICATION: 435
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: 08/133,530
CC FILING DATE: 07-OCT-1993
CC ATTORNEY/AGENT INFORMATION:
CC NAME: Mack, Susan J.
CC REGISTRATION NUMBER: 30,951
CC REFERENCE/DOCKET NUMBER: A6462
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: (202)293-7060
CC TELEFAX: (202)293-2920
CC TELEX: 6491103
CC INFORMATION FOR SEQ ID NO: 18:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 395 amino acids
CC TYPE: amino acid
CC TOPOLOGY: linear
CC MOLECULE TYPE: protein
CC SEQUENCE 395 AA; 42522 MW; 753192 CN;

Query Match 6.7%; Score 107; DB 1; Length 395;
Best Local Similarity 29.7%; Pred. No. 3.56e+00;
Matches 27; Conservative 25; Mismatches 33; Indels 6; Gaps 5;

Db 180 ELTPENAEAVARFLGDAVDREP-ALMLYFECRCAREESKRVPPRTFGSAPRLTDEDFGLL 238
Qy 28 QLPPEVA-AMARLLGD-LDRSTFRLLKRVVSSLOGEDCRDGVORLGSANLPEQLGAL 85
Db 239 NTALAEARRLCIDLPVP--PNAYPYHLRE 267
Qy 86 LAGMHTLLQALRLPPTSILKPDTPFRD-QLQE 115

RESULT 7
ID US-08-576-626A-31 STANDARD; PRT; 1114 AA.
XX AC xxxxxx
XX DT
XX DE
XX SEQUENCE 31, Application US/08576626A
XX SEQUENCE 31, Application US/08576626A
XX Patent No. 5998194
XX GENERAL INFORMATION:
XX APPLICANT: Summers, R.G.
XX APPLICANT: Katz, L.
XX APPLICANT: Donadio, S.
XX APPLICANT: Staver, M.J.
XX TITLE OF INVENTION: POLYKETIDE-ASSOCIATED SUGAR
XX NUMBER OF SEQUENCES: 60
XX CORRESPONDENCE ADDRESSES:
XX ADDRESSEE: Abbott Laboratories
XX STREET: 100 Abbott Park Road
XX CITY: Abbott Park
XX STATE: Illinois
XX COUNTRY: USA
XX ZIP: 60064-3500
XX COMPUTER READABLE FORM:
XX MEDIUM TYPE: Diskette
XX COMPUTER: IBM Compatible
```

```
CC OPERATING SYSTEM: DOS
CC SOFTWARE: FastSeq Version 2.0
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/08/576,626A
CC FILING DATE: 21-DEC-1995
CC CLASSIFICATION: 435
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER:
CC FILING DATE:
CC ATTORNEY/AGENT INFORMATION:
CC NAME: Dianne Casuto
CC REGISTRATION NUMBER: P-40,943
CC REFERENCE/DOCKET NUMBER: 5857.US.O1
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: (847) 938-3137
CC TELEFAX: (847) 938-2623
CC TELEX:
CC INFORMATION FOR SEQ ID NO: 31:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 1114 amino acids
CC TYPE: amino acid
CC STRANDEDNESS: single
CC TOPOLOGY: linear
CC MOLECULE TYPE: No. 5998194e
CC SEQUENCE 1114 AA; 120173 MW; 5916545 CN;

Query Match 6.7%; Score 107; DB 2; Length 1114;
Best Local Similarity 25.2%; Pred. No. 3.56e+00;
Matches 33; Conservative 36; Mismatches 54; Indels 8; Gaps 7;

Db 150 MTAVLGAAGVLR--GAANDARVS-LDAQLSQQAVTEAAVAALPADPALRAL-FAGAEM 205
Qy 1 MSAVGAATPYLHHPGDSHSRVSFLGAQLPPE-VAAMARLLGLDRSTFRKLLKFWVSSL 59
Db 206 TANTVDAVLAVSREPGLAE-RIADDPAAAOQTVAEVLRLHPA-LHLER-RTATAEVRLG 262
Qy 60 QGEDCRDGVORLGSANLPEEQLGALLAGMHTLLQALRLPPTSILKPDTPFRDQLQELCIP 119
Db 263 EHVIAGEGEV 273
Qy 120 QDLVGDLASVV 130

RESULT 8
ID US-09-015-815-1 STANDARD; PRT; 448 AA.
XX AC xxxxxx
XX DT
XX DE
XX SEQUENCE 1, Application US/09015815
XX SEQUENCE 1, Application US/09015815
XX Patent No. 5965356
XX GENERAL INFORMATION:
XX APPLICANT: AURELIAN, LAURE
XX APPLICANT: SMITH, CYNTHIA
XX TITLE OF INVENTION: HERPES SIMPLEX VIRUS TYPE SPECIFIC SEROASSAY
XX FILE REFERENCE: 1437LA
XX CURRENT APPLICATION NUMBER: US/09/015,815
XX CURRENT FILING DATE: 1998-01-29
XX EARLIER APPLICATION NUMBER: US 60/036,622
XX EARLIER FILING DATE: 1997-01-31
XX NUMBER OF SEQ ID NOS: 1
XX SOFTWARE: PatentIn Ver. 2.0
XX SEQ ID NO: 1
XX LENGTH: 448
XX TYPE: PRT
XX ORGANISM: herpes simplex virus-2
XX FEATURE:
XX NAME/KEY: DOMAIN
XX LOCATION: (1)..(446)
XX OTHER INFORMATION: PROTEIN KINASE DOMAIN OF ICP10 SUBUNIT OF HSV-2
```


Query Match 6.1%; Score 98; DB 3; Length 1082;
Best Local Similarity 22.8%; Pred. No. 1.43e+01;
Matches 18; Conservative 26; Mismatches 29; Indels 6; Gaps 5;
Db 19 LSTEGALHVLLEPARGPPQRL-SFSFGDHLAEDLCVQAKASAILPV-Y--H-SLFAL 73
QY 82 LGALLAGMHTLLQOALRLPPTSLKPTDFRDOL-OELCIPQDLYGDLASVFGSQRELLDS 140
Db 74 ATEDLSCWFPFRATSPWRM 92
QY 141 VAOQOGANLPHVADFRRV 159

Search completed: Fri May 12 10:36:40 2000
Job time : 22 secs.

WORLD (TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

Mpsrch_pp protein - protein database search, using Smith-Waterman algorithm
Run on: Fri May 12 10:35:27 2000; MasPar time 29.93 Seconds
Tabular output not generated. 177.250 Million cell updates/sec

Title: >US-09-223-796-4
Description: (1-224) from US09223796.pep
Perfect Score: 1604
Sequence: 1 MSVAGAAATPYLHHFGDSHG.....ALVLKEMADLEKRCERRLOD 224

Scoring table: PAM 150
Gap 11

Searched: 188963 seqs, 23686106 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: a-geneseq36
1:geneseqp

Statistics: Mean 33.179; Variance 157.131; scale 0.211

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description	Pred. No.
1	1319	82.2	224	1	W37723 Rat Hypertension relat	1.42e-104
2	122	7.6	195	1	W71584 Amino acid sequence of	6.45e-01
3	122	7.6	196	1	Y07867 Human secreted protein	6.45e-01
4	116	7.2	169	1	W88390 Human zneul partial po	1.72e+00
5	116	7.2	181	1	W88391 Human zneul partial po	1.72e+00
6	116	7.2	234	1	W88382 Human neuro-growth fac	1.72e+00
7	116	7.2	273	1	W88381 Human neuro-growth fac	1.72e+00
8	110	6.9	800	1	R84883 Transcription factor p	4.50e+00
9	110	6.9	2414	1	R84882 Transcription factor p	4.50e+00
10	110	6.9	2414	1	W40057 Cellular transcription	4.50e+00
11	107	6.7	360	1	W19734 Sugar biosynthesis enz	7.24e+00
12	107	6.7	361	1	W99387 S.erythraea dtpd-4-ket	7.24e+00
13	106	6.6	339	1	R77317 Protein activated lipa	8.48e+00
14	106	6.6	727	1	W99798 Human VRRP-1 (VR2) cap	8.48e+00
15	104	6.5	344	1	R47213 Lipase modulator.	1.16e+01
16	100	6.2	694	1	R04107 DNA-binding protein GC	2.16e+01
17	100	6.2	784	1	W34179 Human GC binding prote	2.16e+01
18	100	6.2	1144	1	W72206 HSV-2 strain SB5 Conti	2.16e+01
19	100	6.2	1180	1	W72020 HSV-2 strain SB5 Conti	2.16e+01
20	100	6.2	1180	1	W72098 HSV-2 strain SB5 Conti	2.16e+01
21	98	6.1	1064	1	R96037 Protein tyrosine kinas	2.94e+01
22	95	5.9	603	1	R85889 WD-40 domain-contg. ra	4.65e+01
23	94	5.9	1137	1	P71182 Sequence of herpes sim	5.41e+01

24	93	5.8	1291	1	R68975	Entamoeba histolytica	6.29e+01
25	93	5.8	1292	1	R70681	Entamoeba histolytica	6.29e+01
26	92	5.7	335	1	R99975	Pseudomonas sp. SD705	7.31e+01
27	92	5.7	350	1	W06119	Human liver activin be	7.31e+01
28	92	5.7	350	1	R92754	Human growth different	7.31e+01
29	91	5.7	476	1	W02616	Wheat adenylsuccinate	8.49e+01
30	91	5.7	490	1	W99454	Arabidopsis adenylsuc	8.49e+01
31	91	5.7	490	1	R97733	Arabidopsis adenylsuc	8.49e+01
32	91	5.7	490	1	W02614	Arabidopsis adenylsuc	8.49e+01
33	91	5.7	822	1	W41943	Human OC-116 kDa prote	8.49e+01
34	91	5.7	1319	1	R45002	Cellulose synthase ope	8.49e+01
35	91	5.7	1326	1	W69758	Acetobacter xylinum bc	8.49e+01
36	91	5.7	1336	1	W64332	Seq ID #8 from BE19653	8.49e+01
37	90	5.6	332	1	W52284	Mutant prenyl diphosph	9.86e+01
38	90	5.6	561	1	W63701	Human hsk1 protein.	9.86e+01
39	89	5.5	182	1	W20289	H. pylori transporter	1.14e+02
40	88	5.5	279	1	W74908	Human secreted protein	1.33e+02
41	89	5.5	494	1	R31888	Defective tyrosine kin	1.14e+02
42	89	5.5	494	1	R76615	Saccharomyces cerevisi	1.14e+02
43	89	5.5	494	1	R56519	Protein kinase (HRR25)	1.14e+02
44	89	5.5	519	1	W20640	H. pylori transporter	1.14e+02
45	88	5.5	686	1	R25591	RING11 antigenic pepti	1.33e+02

ALIGNMENTS

RESULT 1
ID W37723 standard; Protein; 224 AA.
AC W37723;
DT 09-JUN-1998 (first entry)
DE Rat Hypertension related calcium regulator.
KW Hypertension related calcium regulated gene; HCARG; rat parathyroid;
KW extracellular calcium concentration; antibody; hypertension;
KW hyperthyroidism; osteoporosis; heart failure; diabetes; stroke;
KW cancer; inflammatory disease; asthma.
OS Rattus rattus.
FH Key Location/Qualifiers
FT Misc_difference 15..21 /note= "EF-hand like motif"
PN WO9749807-A2.
PD 31-DEC-1997.
PF 23-JUN-1997; CA0439.
PR 21-JUN-1996; US-867495.
PA (GOSS/) GOSSARD F.
PA (HAME/) HAMEET P.
PA (LEWA/) LEWANCZUK R.
PA (TREW/) TREMBLAY J.
PI Gossard F, Hamet P, Lewanczuk R, Tremblay J;
DR WPI; 98-077171/07.
DR N-PSDB; V18890.

PT Hypertension related calcium regulated gene - useful to develop products to treat or detect, e.g. hypertension, stroke, osteoporosis, heart failure, cancer, diabetes or asthma
PS Claim 8; Pages 26-27; 46pp; English.
CC This is the amino acid sequence of the hypertension related calcium regulated gene (HCARG), which was isolated from the rat parathyroid. Its expression is regulated by extracellular calcium concentration. An antibody against the protein, can be used to detect or modulate (e.g. enhance or inhibit) abnormal calcium levels. They can specifically be used to detect or treat, e.g. hypertension, hyperthyroidism, osteoporosis, heart failure, diabetes, cancer, inflammatory disease, and asthma.
SQ Sequence 224 AA;

Query Match 82.2%; Score 1319; DB 1; Length 224;
Best Local Similarity 79.5%; Pred. No. 1.42e-104;
Matches 178; Conservative 33; Mismatches 13; Indels 0; Gaps 0;
Db 1 MSALGAAPYLHHPADSHGRVSFLGSSQSPFVTAVALKDLDRSTFRKLKLVVGAH 60
QY 1 MSVAGAAATPYLHHFGDSHGSRVSFLGQAQLPPEVAARLLGDLDRSTFRKLKLVVSSIQ 60
Db 61 GKDCREAVQLGASANLSERLAVLLAGHTTLLQALRLFPASPKPDAFQEELGILGIPQ 120

QY 61 GEDCRGVQRUGV SANUPPEQLGALLAGMFTLQOALRLPPTSLKPTDFRDLQELCIPQ 120
DB 121 DLIGDLASLAFSGRPLDLSVAQOQSGSLPHVSFVRWVDVAISTSAQSRSLQPSVLMLQ 180
QY 121 DLVGLASVFGSRPLDLSVAQOQGWLPVADFRWVDVAISTSAQSRSLQPSVLMLQ 180
DB 181 KLDTGSARHFEVPIAKFQELRYSLVALVKEMAELEKCKERKLOD 224
QY 181 KLSGSAFYFEVPTAKFQELRYSLVALVKEMAELEKCKERKLOD 224

RESULT 2

ID W1684 standard; Protein; 195 AA.
AC W1684;
DT 04-DEC-1998 (first entry)
DE Amino acid sequence of the human tumorigenesis associated protein.
KW Human; tumorigenesis associated protein; HTAP; transplantation;
KW tumour; Antagonist; cancer; inflammation; immunological disease;
KW antibody; Probe; primer; PCR; amplification; hybridisation;
KW inhibition.
OS Homo sapiens.
PN WO9841635-A1.
PD 24-SEP-1998.
PF 20-MAR-1998; U06066.
PR 20-MAR-1997; US-822260.
PA (INCY-) INCYTE PHARM INC.
PI Goli SK, Hillman JL;
DR WPI; 98-521224/44.
DR N-PSDB; V58281.
PT New tumorigenesis-associated protein and related nucleic acid,
PT vectors, transformed cells - antibodies, agonists and antagonists,
PT for diagnosis, treatment and prevention of abnormal cellular
PT differentiation, particularly cancers and inflammation
PS Claim 1; Fig 1A-1B; 54pp; English.
CC This is the amino acid sequence of the human tumorigenesis
CC associated protein (HTAP), used in the method of the invention. HTAP,
CC is involved with cell proliferation and inflammation. It can be used
CC to stimulate cell proliferation (e.g. of cells intended for
CC transplantation in treatment of tumours or infections, or to treat
CC genetic defects). Antagonists of HTAP are used to treat or prevent
CC a wide range of cancers (adenocarcinoma, melanoma, sarcoma, lymphoma,
CC leukaemia etc.), also inflammation where associated with infection or
CC immunological disease (e.g. asthma, cystic fibrosis, rheumatoid
CC arthritis). HTAP is also used to raise antibodies and to screen
CC libraries for specific-binding agents. The antibodies are used as for
CC diagnosis or monitoring of HTAP-related diseases (in usual
CC immunoassays), in competitive drug screens and to isolate HTAP from
CC its natural sources. HTAP derived probes or primers, are used in
CC standard amplification or hybridisation tests to diagnose HTAP-related
CC diseases; to identify related sequences; for genomic mapping and for
CC screening for specific inhibitors.
SQ Sequence 195 AA;

Query Match 7.6%; Score 122; DB 1; Length 195;
Best Local Similarity 24.8%; Pred. No. 6.45e-01;
Matches 27; Conservative 35; Mismatches 38; Indels 9; Gaps 8;
DB 87 LED-CKFRDRI-ELFCTEQNNKNSLEILLGSGRPLDLSVAQOQGWLPVADFRWVDVAISTSAQSRSL 144
QY 113 LQELC-IPQDLVGLASVFGSRPLDLSVAQOQGWLPVADFRWVDVAISTSAQSRSL 171
DB 145 YRPAYLVTLVSQNTDPSYP-EISFSCSMEQLQDLVG-KLKDASKSLER 191
QY 172 LQPSVLMLQKL--SDGSAYRFEVP-TAKFQELRYSLVALVKEMA-DLEK 216

RESULT 3

ID Y07867 standard; Protein; 196 AA.
AC Y07867;
DT 06-JUL-1999 (first entry)
DE Human secreted protein fragment encoded from gene 16.
KW Human; secreted protein; treatment; prevention; protein therapy; AIDS;

KW gene therapy; diagnosis; cancer; tumour; neurodegenerative disorder;
KW developmental abnormality; fetal deficiency; blood disorder; leukemia;
KW immune system disease; autoimmune disease; hepatic disease; lymphoma;
KW renal disease; inflammation; allergy; Alzheimer's disease; schizophrenia;
KW cognitive disorder; prostate disease; skeletal; cardiac; muscle disorder;
KW pulmonary disorder; transplant rejection; osteoclast; osteoporosis;
KW arthritis; malignancy; digestive; endocrine; infection.
OS Homo sapiens.
PN WO9918208-A1.
PD 15-APR-1999.
PF 01-OCT-1998; U20775.
PR 02-OCT-1997; US-060884.
PR 02-OCT-1997; US-060833.
PR 02-OCT-1997; US-060836.
PR 02-OCT-1997; US-060837.
PR 02-OCT-1997; US-060838.
PR 02-OCT-1997; US-060839.
PR 02-OCT-1997; US-060843.
PR 02-OCT-1997; US-060862.
PR 02-OCT-1997; US-060866.
PR 02-OCT-1997; US-060874.
PA (HUMA-) HUMAN GENOME SCI INC.
PI Carter KC, Duan DR, Endress GA, Feng P, Ferrie AM,
PI Florence KA, Greene JM, Janat F, Lafleur DW, Ni J,
PI Rosen CA, Ruben SM, Shi Y, Young P, Yu G;
DR WPI; 99-264022/22.
DR N-PSDB; X37466.
PT New isolated human genes and the secreted polypeptides they encode
PS Claim 1b; Page 285-286; 368pp; English.
CC This invention describes novel isolated human genes and the secreted
CC proteins they encode. The products of the invention are useful for
CC preventing, treating or ameliorating medical conditions, e.g. by protein
CC or gene therapy. Also pathological conditions can be diagnosed by
CC determining the amount of the new polypeptides in a sample or by
CC determining the presence of mutations in the new polynucleotides.
CC Specific uses are described for each of the 101 polynucleotides, based on
CC which tissues they are most highly expressed in, and include developing
CC products for the diagnosis or treatment of cancer, tumours,
CC neurodegenerative disorders, developmental abnormalities and fetal
CC deficiencies, blood disorders, leukemias, diseases of the immune system,
CC autoimmune diseases, hepatic and renal disease, lymphomas, inflammation,
CC allergies, Alzheimer's and cognitive disorders, schizophrenia, prostate
CC disease, skeletal or cardiac muscle disorders, pulmonary disorders,
CC transplant rejection, disorders involving osteoclasts such as
CC osteoporosis, arthritis or malignancies, digestive/endocrine disorders,
CC infections and AIDS. The human secreted proteins of the invention are
CC represented in Y07852-Y07993 and the encoding nucleic acids are
CC represented in X37451-X37552.
SQ Sequence 196 AA;

Query Match 7.6%; Score 122; DB 1; Length 196;
Best Local Similarity 24.8%; Pred. No. 6.45e-01;
Matches 27; Conservative 35; Mismatches 38; Indels 9; Gaps 8;
DB 87 LED-CKFRDRI-ELFCTEQNNKNSLEILLGSGRPLDLSVAQOQGWLPVADFRWVDVAISTSAQSRSL 144
QY 113 LQELC-IPQDLVGLASVFGSRPLDLSVAQOQGWLPVADFRWVDVAISTSAQSRSL 171
DB 145 YRPAYLVTLVSQNTDPSYP-EISFSCSMEQLQDLVG-KLKDASKSLER 191
QY 172 LQPSVLMLQKL--SDGSAYRFEVP-TAKFQELRYSLVALVKEMA-DLEK 216

RESULT 4

ID W88390 standard; Protein; 169 AA.
AC W88390;
DT 26-APR-1999 (first entry)
DE Human Znu1 partial polypeptide.
KW Znu1; neuro-growth factor-like protein; human; breast cancer;
KW glioblastoma; pituitary adenoma; Alzheimer's disease; therapy;
KW nerve regeneration; haematopoiesis; fertility; contraception;
KW antibody.
OS Homo sapiens.

OY 65 RDGVORLGSANLPPEQLGALLAGMHTLLOQAL-R-LP-PTSLKPDPTFRDQL 113

RESULT 7

ID W88381; standard; Protein; 273 AA.
AC W88381;
DE Transcription factor p300 C-terminal (1572-2371) region.
KW Transcription factor; p300; adenovirus; early region 1A; E1A;
KW cancer; diagnosis; monoclonal antibody.
OS Homo sapiens.
PN W09528499-A1.
PD 26-OCT-1995.
PF 13-APR-1995; U04682.
PR 14-APR-1994; US-227536.
PA (DAND) DANA FARBER CANCER INST INC.
PI Eckner R, Ewen M, Livingston D;
DR WPI: 95-373813/48.
PT Nucleic acid encoding human p300 that associates with adenovirus E1A
FT - and related vectors, host cells and screening assays, also
FT diagnosis of cancerous and pre-cancerous tissue by detection of
FT mutant p300
FT Claim 17; Page 72-78; 126pp; English.
PS A fusion protein comprising glutathione-S-transferase and amino
FT acids 1572-2371 (R84883) of human transcription factor p300
CC (full sequence given in R84882) was used to produce an anti-p300
CC monoclonal antibody (Mab) useful in identifying p300 binding
CC proteins and for characterizing p300 in immunoassays.
SQ Sequence 800 AA;

Query Match 7.2%; Score 116; DB 1; Length 273;
Best Local Similarity 44.2%; Pred. No. 1.72e+00;
Matches 23; Conservative 9; Mismatches 16; Indels 4; Gaps 4;

DB 197 KEQVORLQSRVLDLEKQLVLAPLHSLASQALEHGLPDPGSLVHSFO-QL 247

OY 65 RDGVORLGSANLPPEQLGALLAGMHTLLOQAL-R-LP-PTSLKPDPTFRDQL 113

RESULT 8

Query Match 7.2%; Score 116; DB 1; Length 273;
Best Local Similarity 44.2%; Pred. No. 1.72e+00;
Matches 23; Conservative 9; Mismatches 16; Indels 4; Gaps 4;

DB 197 KEQVORLQSRVLDLEKQLVLAPLHSLASQALEHGLPDPGSLVHSFO-QL 247

OY 65 RDGVORLGSANLPPEQLGALLAGMHTLLOQAL-R-LP-PTSLKPDPTFRDQL 113

ID R84883 standard; Protein; 800 AA.
AC R84883;
DE Transcription factor p300 C-terminal (1572-2371) region.
KW Transcription factor; p300; adenovirus; early region 1A; E1A;
KW cancer; diagnosis; monoclonal antibody.
OS Homo sapiens.
PN W09528499-A1.
PD 26-OCT-1995.
PF 13-APR-1995; U04682.
PR 14-APR-1994; US-227536.
PA (DAND) DANA FARBER CANCER INST INC.
PI Eckner R, Ewen M, Livingston D;
DR WPI: 95-373813/48.
PT Nucleic acid encoding human p300 that associates with adenovirus E1A
FT - and related vectors, host cells and screening assays, also
FT diagnosis of cancerous and pre-cancerous tissue by detection of
FT mutant p300
FT Claim 17; Page 72-78; 126pp; English.
PS A fusion protein comprising glutathione-S-transferase and amino
FT acids 1572-2371 (R84883) of human transcription factor p300
CC (full sequence given in R84882) was used to produce an anti-p300
CC monoclonal antibody (Mab) useful in identifying p300 binding
CC proteins and for characterizing p300 in immunoassays.
SQ Sequence 800 AA;

Query Match 6.9%; Score 110; DB 1; Length 800;
Best Local Similarity 34.0%; Pred. No. 4.50e+00;
Matches 17; Conservative 13; Mismatches 18; Indels 2; Gaps 2;

DB 566 QAGVORAGLPQQOQQOQPPMGSMQAQQ-MNMHNHTM-PSQFRDILR 613

OY 65 RDGVORLGSANLPPEQLGALLAGMHTLLOQALRPPPTSLKPDPTFRDQL 114

RESULT 9

ID R84882 standard; Protein; 2414 AA.
AC R84882;
DE Transcription factor p300.
KW Transcription factor; p300; adenovirus; early region 1A; E1A;
KW cancer; diagnosis.
OS Homo sapiens.
FH Key Location/Qualifiers
FT Peptide 11..17
FT region /label= Nuclear_location_signal
FT 342..421
FT /label= C/H-rich_region_1
FT /note= "cysteine/histidine-rich region containing
FT 2 putative zinc finger motifs"
FT domain 1070..1134
FT /label= Bromodomain
FT region 1162..1461
FT /label= C/H-rich_region_2
FT region 1622..1821
FT /label= C/H-rich_region_3
PN W09528499-A1.
PD 26-OCT-1995.
PF 13-APR-1995; U04682.
PR 14-APR-1994; US-227536.
PA (DAND) DANA FARBER CANCER INST INC.
PI Eckner R, Ewen M, Livingston D;
DR WPI: 95-373813/48.
PT Nucleic acid encoding human p300 that associates with adenovirus E1A
FT - and related vectors, host cells and screening assays, also
FT diagnosis of cancerous and pre-cancerous tissue by detection of
FT mutant p300
PS Disclosure; Page 61-78; 126pp; English.
CC Transcription factor p300 was isolated from human 293 cells and used
CC to generate polyclonal antisera in mice. These were used to screen
CC 293 cDNA libraries to isolate clones contg. overlapping inserts,
CC which were assembled to obtain a full-length cDNA sequence (T02792)

RESULT 9

CC encoding a protein (R84882) of predicted mol.wt. 264.236 kDa.
CC p300 may be produced in host cells (pref. mammalian) and used to
CC raise monoclonal antibodies or to screen cpds. for the ability to
CC modulate p300-dependent transcription.
SQ Sequence 2414 AA;

Query Match 6.9%; Score 110; DB 1; Length 2414;
Best Local Similarity 34.0%; Pred. No. 4.50e+00;
Matches 17; Conservative 13; Mismatches 18; Indels 2; Gaps 2;

Db 2137 QAGVQRAGLPQQPQQQLQPMGMSPPAQQ-MNMNHTM-PSQFRDLIR 2184
QY 65 RDGVQRGLVSNLPEQLGALLAGMHTLLQALRLPPTSLKPDTRDQLQ 114

RESULT 10
ID W40057 standard; Protein; 2414 AA.
AC W40057;
DT 20-JUL-1998 (first entry)
DE Cellular transcriptional factor p300.
KW Cellular transcriptional factor; p300; human; p300; P/CAF;
KW transcription; histone acetyltransferase; HIV; infection; cancer;
KW therapy; muscle differentiation.
OS Homo sapiens.
FH Key Location/Qualifiers
FT Region 1763..1966
FT /note= "P/CAF binding region"

PN W0903652-A2.
PD 29-JAN-1998.
PF 23-JUL-1997; U12877.
PR (USSH) US DEPT HEALTH & HUMAN SERVICES.
PI Howard BH, Nakatani Y;
DR WPI: 98-12077/11.
DR N-PSDB: V10092.
PT New isolated p300/CBP-associated factor, P/CAF - used to develop
PT products for modulating transcription, e.g. for treating HIV
PT infection or cancers or for promoting muscle differentiation
PS Disclosure; Page 76-81; 107pp; English.
CC This polypeptide sequence comprises p300, a global transcriptional
CC coactivator that is involved in the regulation of various
CC DNA-binding transcriptional factors. The invention relates to a
CC novel human p300/CBP associated cofactor, P/CAF (see W40052), that
CC modulates transcription through binding to p300 and CBP (see
CC W40058). The region (see W40055) of p300 that binds to P/CAF
CC is claimed. The invention provides methods of screening for
CC compounds that inhibit or stimulate the transcription modulating
CC and histone acetyltransferase activity of P/CAF and p300/CBP.
CC Inhibitors can be used e.g. to inhibit HIV TAR-mediated
CC transcription in the treatment of HIV infection. Stimulators can
CC be used e.g. to activate tumour suppressor p53 in the treatment of
CC cancer or to activate the muscle differentiation factor MyoD to
CC promote muscle differentiation. The products can also be used to
CC inhibit the cell cycle progression inducing effect of an
CC oncoprotein which binds p300/CBP in a subject. Also provided is
CC a method for determining the amount of P/CAF in a sample by
CC contacting the sample with the P/CAF binding region of p300 and
CC determining the amount of P/CAF/p300 complex formed.
SQ Sequence 2414 AA;

Query Match 6.9%; Score 110; DB 1; Length 2414;
Best Local Similarity 34.0%; Pred. No. 4.50e+00;
Matches 17; Conservative 13; Mismatches 18; Indels 2; Gaps 2;

Db 2137 QAGVQRAGLPQQPQQQLQPMGMSPPAQQ-MNMNHTM-PSQFRDLIR 2184
QY 65 RDGVQRGLVSNLPEQLGALLAGMHTLLQALRLPPTSLKPDTRDQLQ 114

RESULT 11

ID W19734 standard; Protein; 360 AA.
AC W19734;
DT 18-SEP-1997 (first entry)

DE Sugar biosynthesis enzyme EryCII.
KW Polyketide; glycosylation; eryCII; D-desosamine; antimicrobial;
KW antibiotic; antifungal; fungicide; anticancer; cytostatic;
KW anthelmintic.
OS Saccharopolyspora erythraea.
PN W09723630-A2.
PD 03-JUL-1997.
PF 23-DEC-1996; U20238.
PR 21-DEC-1995; US-576626.
PA (ABBO) ABBOTT LAB.
PI Donadio S, Katz L, Staver MJ, Summers RG;
DR WPI: 97-351066/32.
DR N-PSDB; T72684.
PT New genes involved in sugar biosynthesis and attachment - used to
PT generate polyketide antimicrobials etc. with altered pattern of
PT glycosylation
PS Disclosure; Fig 4A; 85pp; English.
CC EryCII (W19734) is an enzyme involved in the biosynthesis of the
CC sugar D-desosamine. It is one of 10 enzymes (see also W19735-43)
CC predicted to be involved in D-desosamine or L-mycarose biosynthesis
CC and attachment that have been identified from gene clusters (T72684
CC and T72685) of Saccharopolyspora erythraea. Novel glycosylation-
CC modified polyketides are produced by selectively altering,
CC inactivating or augmenting the eryA and/or eryC genes encoding
CC these sugar biosynthesis enzymes and introducing them into
CC polyketide-producing microorganisms.
SQ Sequence 360 AA;

Query Match 6.7%; Score 107; DB 1; Length 360;
Best Local Similarity 25.2%; Pred. No. 7.24e+00;
Matches 33; Conservative 36; Mismatches 54; Indels 8; Gaps 7;
Db 150 MTAVLGAAGVLR--GAAMDARVS-LDAQLSPQOLAVTEAAVAALPADPALRAL-FAGAEM 205
QY 1 MSVAGATPTLHPHSGSRVSFLGQLPPE-VAAMARLLGLDSTFKLKFVSSL 59
Db 206 TANTVVDAVLAVSAEPGLAE-RIADDPAAQRTVAEVLRLHPA-LHLER-RTATAEVLRL 262
QY 60 QGEDCDGVRQLGVSANLPEELGALLAGMHTLLQALRLPPTSLKPDTRDQLQELCIP 119
Db 263 EHVIGEGEVV 273
QY 120 QDLVGDLSVV 130

RESULT 12
ID W99387 standard; Protein; 361 AA.
AC W99387;
DT 08-JUN-1999 (first entry)
DE S.erythraea dtdp-4-keto-L-6-deoxyhexose-3,4-isomerase.
KW Gene cluster; bacterium; enzyme; macrolide; antibiotic; erythromycin;
KW secondary metabolite; eryBII; eryCII; hybridisation; probe;
KW glycosylation; macrolactone; oleandomycin.
OS Saccharopolyspora erythraea.
PN W09905283-A2.
PD 04-FEB-1999.
PF 21-JUL-1998; F01593.
PR 12-JUN-1998; FR-007411.
PR 25-JUL-1997; FR-009458.
PA (HMRI) HOSCHST MARION ROUSSEL.
PI Cortes J, Gaisser S, Leadlay P, Michel JM, Raynal MC,
PI Salah-Bey K, Fromentin C, Mendez C, Salas JA;
DR WPI: 99-142938/12.
DR N-PSDB: X25772.
PT New nucleic acid sequences encoding enzymes involved in macrolide
PT biosynthesis - useful for producing hybrid secondary metabolites,
PT particularly erythromycin analogues
PS Claim 6; Fig 2; 221pp; French.
CC This sequence represents the dtdp-4-keto-L-6-deoxyhexose-3,4-isomerase
CC enzyme encoded by the eryCII gene from the eryG-eryAIII gene cluster from
CC the Gram-positive bacterium Saccharopolyspora erythraea. This gene
CC cluster encodes enzymes involved in the production of the macrolide
CC antibiotic erythromycin as a secondary metabolite. The erythromycin gene

CC cluster spans approximately 53 kb and contains at least 20 open reading
 CC frames (ORF). The genes are used to produce hybrid secondary metabolites
 CC in *S. erythraea*, i.e. erythromycin analogues which may have improved
 CC properties or as hybridisation probes for isolating homologous genes
 CC involved in glycosylation of macrolactones in macrolide-producing
 CC strains (specifically oleandomycin-producing strains of *Streptomyces*
 CC antibioticus).
 SQ Sequence 361 AA;

Query Match 6.7%; Score 107; DB 1; Length 361;
 Best Local Similarity 25.2%; Pred. No. 7.24e+00;
 Matches 33; Conservative 36; Mismatches 54; Indels 8; Gaps 7;
 Db 151 MTAVLGAAGVLR--GAANDARYS-LDAQLSPQQLAVTEAAVAALPADPALRAL-PAGAEM 206
 QY 1 MSAVGAATPYLHHPGDSHGRVSFLGAQLPPE-VAAMARLLGDLDRSFRKLLKFFVSSL 59
 Db 207 TANTVVDAVLAVSAPGLAE-RIADDPAAQRTVAEVLRLHPA-LHLER-RTATAEVRLG 263
 QY 60 QGEDCRDGVQRLGVSANLPEQLGALLAGMHTLLQALRLPPTSLKPDTRDQLQELCIP 119
 Db 264 EHVIAGEBEV 274
 QY 120 QDLVGDLSAV 130

RESULT 13
 ID R77317 standard; Protein; 339 AA.
 AC R77317;
 DT 07-FEB-1996 (first entry)
 DE Protein activated lipase B.
 KW Lipase B; activate; glutathionine transferase.
 OS Pseudomonas aeruginosa.
 PN J07135971-A.
 PD 30-MAY-1995.
 PF 05-JUL-1993; 165561.
 PR 05-JUL-1993; JP-165561.
 PA (TOYM) TOYORO KK.
 DR WPI; 95-227395/30.
 DR N-PSDB; 091399.
 PT A new protein for activating lipase - may be produced as a fusion
 PT protein with glutathione transferase, used in the production of
 PT lipase
 PS Claim 1; Page 6-7; 9pp; Japanese.
 CC R77317 represents a new protein-activated lipase. The protein is
 CC produced as a fusion protein with glutathionine transferase. The
 CC protein has an amino acid sequence of 339 residues and a molecular
 CC weight of 39 kD. The method used to produce the new protein yields
 CC active lipase of very high purity without using any complex
 CC purification process such as chromatography.
 SQ Sequence 339 AA;

Query Match 6.6%; Score 106; DB 1; Length 339;
 Best Local Similarity 30.5%; Pred. No. 8.48e+00;
 Matches 18; Conservative 21; Mismatches 17; Indels 3; Gaps 2;

Db 207 LSTEEKAAIDRL--RASLPDQESVLPQLQSELOQQTALQAGAGPEAIRMQROL 263
 QY 59 LOGEDCRDGVQRLGVSANLPEQLGALLAGMHT-LLQALRLPPTSLKPDTRDQLQEL 116

RESULT 14
 ID W99798 standard; Protein; 727 AA.
 AC W99798;
 DT 16-JUN-1999 (first entry)
 DE Human VRRP-1 (VR2) capsaicin receptor.
 KW VRI; capsaicin receptor; VR2; VRRP-1; analgesic; diagnosis;
 KW human disease; painful syndrome.
 OS Homo sapiens.
 PN W09909140-A1.
 PD 25-FEB-1999.
 PF 20-AUG-1998; U17466.
 PR 22-JAN-1998; US-072151.

PR 20-AUG-1997; US-915461.
 PA (REGC) UNIV CALIFORNIA.
 PI Brake A, Caterina M, Julius DJ;
 DR WPI; 99-181023/15.
 FT New capsaicin receptor polypeptide - useful for screening or
 FT characterising capsaicin receptor-binding compounds
 PS Claim 4; Page 86-88; 99pp; English.
 CC The present sequence is an isolated capsaicin receptor polypeptide
 CC (1). Capsaicin polypeptides are useful for identifying binding compounds
 CC which affect cellular responses. Preferably this is for identifying a
 CC compound that binds (1) and affects a cellular response associated with
 CC capsaicin biological activity (e.g. intracellular calcium flux). The
 CC polypeptides and host cells are useful for detecting a vanilloid
 CC compound (an essential structural component of capsaicin) from natural
 CC products by detecting an alteration of intracellular response associated
 CC with capsaicin receptor activity, preferably an alteration of
 CC intracellular calcium levels, and are useful for screening for compounds
 CC for use in analgesics. Capsaicin receptor polypeptides and antibodies
 CC are useful for diagnosis and treatment of human diseases and painful
 CC syndromes. The transgenic mammals can be used to screen for capsaicin
 CC receptor antagonists and agonists. Prior art methods for screening or
 CC characterising new capsaicin receptor-binding compounds relied on assays
 CC using sensory neurons in culture or in intact animals. The new
 CC polypeptides provide a more sensitive screen.
 SQ Sequence 727 AA;

Query Match 6.6%; Score 106; DB 1; Length 727;
 Best Local Similarity 29.3%; Pred. No. 8.48e+00;
 Matches 24; Conservative 24; Mismatches 26; Indels 8; Gaps 8;
 Db 236 ASLQATDSQGNVYLHALVMISDMSAENI-ALVTSMYDGLLQAGARLQCT-VQLEDIRN-L 292
 QY 57 SSLQGEDCR-DGV-QRLGVSANLPEQLGALLAGMHT-LLQALRLPPTSLKPDTRDQL 113
 Db 293 QDL-TPLKLAKEGKIXIFXRH 313
 QY 114 QELCIPQDLVGDLSAV-VFGSQ 134

RESULT 15
 ID R47213 standard; Protein; 344 AA.
 AC R47213;
 DT 27-JUL-1994 (first entry)
 DE Lipase modulator.
 KW Pseudomonas; cloning; lipase gene; lipase modulator gene; lipids;
 KW oils; fats; detergents; diagnostic reagents.
 OS Pseudomonas pseudoalcaligenes.
 PN W09402617-A.
 PD 03-FEB-1994.
 PF 23-JUL-1993; E01995.
 PR (KONN) GIST-BROCADES NV.
 PI Cox MMJ, Gerritse G, Quax WJ;
 DR WPI; 94-048875/06.
 DR N-PSDB; Q56106.
 PT Increasing lipase prodn. in *Pseudomonas* species - by cloning a
 PT lipase gene and a lipase modulator gene into a homologous
 PT *Pseudomonas* species
 PS Disclosure; Page 23; 44pp; English.
 CC The lipase gene from *Pseudomonas pseudoalcaligenes* M1 was cloned in
 CC *E. coli* as in EP-334462. The DNA sequence of insert PvuI/EcoRI
 CC contains two open reading frames, one encoding the lipase gene, the
 CC other encoding a putative lipase modulator gene which increases the
 CC lipase productivity in a homologous host cell. The lipase obtd. is
 CC used for hydrolysing lipids in e.g. fats and oil processing,
 CC detergents or diagnostic reagents.
 CC See also R47212.
 SQ Sequence 344 AA;

Query Match 6.5%; Score 104; DB 1; Length 344;
 Best Local Similarity 21.7%; Pred. No. 1.16e+01;
 Matches 26; Conservative 38; Mismatches 48; Indels 8; Gaps 5;

W P S R L A

(TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MPSrch_pp protein - protein database search, using Smith-Waterman algorithm
Run on: Fri May 12 10:24:14 2000; MasPar time 81.94 Seconds
Tabular output not generated. 189.543 Million cell updates/sec

Title: >US-09-223-796-2
Description: (1-224) from US09223796.pep
Sequence: 1 MSALGAAAPYLHPADSHSG.....ADVLRKEMAEKKCKERKLQD 224

Scoring table: PAM 150
Gap 11

Searched: 225878 seqs, 69334122 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: sptrembl12
1:sp_archaea 2:sp_bacteria 3:sp_fungi 4:sp_human
5:sp_invertebrate 6:sp_mammal 7:sp_mhc 8:sp_organelle
9:sp_phase 10:sp_plant 11:sp_rodent 12:sp_unclassified
13:sp_vertebrate 14:sp_virus

Statistics: Mean 45.804; Variance 101.760; scale 0.450

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
1	124	8.0	195	11	BUP-5'OF BMI-1 PROVIRA	4.75e-03
2	121	7.8	491	2	HYPOTHETICAL 55.6 KD P	1.15e-02
3	119	7.7	202	4	PTD002.	2.05e-02
4	120	7.7	649	4	HEAT SHOCK PROTEIN 75.	1.54e-02
5	118	7.6	512	10	F5A8.3 PROTEIN.	2.74e-02
6	118	7.6	676	5	NEX-2 PROTEIN.	2.74e-02
7	110	7.1	706	2	HYPOTHETICAL 79.8 KD P	2.64e-01
8	107	6.9	470	5	FI3G11.1 PROTEIN.	6.02e-01
9	107	6.9	913	5	CODED FOR BY C. ELEGAN	6.02e-01
10	105	6.8	279	2	FLAGELLIN.	1.03e+00
11	104	6.7	226	14	MC013L.	1.35e+00
12	104	6.7	228	1	228AA LONG HYPOTHETICA	1.35e+00
13	104	6.7	302	9	P2 J HOMOLOG.	1.35e+00
14	104	6.7	382	2	PROBABLE ACYL-COA DEHY	1.35e+00
15	104	6.7	603	11	INSULIN-LIKE GROWTH FA	1.35e+00
16	104	6.7	1116	2	DNA POLYMERASE III SUB	1.35e+00
17	104	6.7	4845	11	UBIQUITIN-CONJUGATING	1.35e+00
18	102	6.6	164	2	GALACTOSE-1-PHOSPHATE	2.30e+00
19	102	6.6	451	2	HYPOTHETICAL 49.7 KD P	2.30e+00
20	103	6.6	555	2	FRUCTOSE-6-P PHOSPHOTR	1.77e+00

21	103	6.6	902	4	043632	SPINDLE POLE BODY PROT	1.77e+00
22	103	6.6	1101	10	Q9ZQT7	L-ZIP+NBS+LRR.	1.77e+00
23	100	6.4	228	2	Q9ZHD3	PUTATIVE TRANSCRIPTION	3.90e+00
24	99	6.4	260	13	Q9M644	PROLIFERATING CELL NUC	5.05e+00
25	99	6.4	266	2	Q47220	FIMBRIAL PROTEIN.	5.05e+00
26	100	6.4	362	1	Q9YF28	362AA LONG HYPOTHETICA	3.90e+00
27	99	6.4	371	2	P74657	UDP-N-ACETYLGLUCOSAMIN	5.05e+00
28	99	6.4	933	3	O14310	NUCLEOPORIN.	5.05e+00
29	99	6.4	1032	3	Q94324	SERINE-THREONINE PROTE	5.05e+00
30	99	6.4	1547	5	Q26471	BLACKJACK.	5.05e+00
31	98	6.3	229	2	P96415	30S RIBOSOMAL PROTEIN	6.54e+00
32	98	6.3	295	5	Q20233	F40F8.7 PROTEIN.	6.54e+00
33	98	6.3	340	2	O84915	GALR-LACI HOMOLOG.	6.54e+00
34	98	6.3	344	2	O53595	HYPOTHETICAL 37.0 KD P	6.54e+00
35	96	6.2	312	2	P72131	PTXR.	1.09e+01
36	96	6.2	312	2	O30374	TRANSCRIPTIONAL ACTIVA	1.09e+01
37	97	6.2	386	2	Q9X8A4	HYPOTHETICAL 40.9 KD P	8.45e+00
38	97	6.2	669	2	Q55421	ELONGATION FACTOR EF-G	1.09e+01
39	96	6.2	960	10	O82318	RECEPTOR-LIKE PROTEIN	1.09e+01
40	96	6.2	1329	5	O76356	C45G7.6 PROTEIN.	1.09e+01
41	96	6.2	1846	5	O61776	F56A6.2 PROTEIN.	1.09e+01
42	96	6.2	2228	10	O48579	HUMAN MI-2 AUTOANTIGEN	1.09e+01
43	95	6.1	287	2	O53288	ABC TRANSPORTER ATP-BI	1.40e+01
44	95	6.1	315	2	O83061	CONSERVED HYPOTHETICAL	1.40e+01
45	95	6.1	452	10	O49368	HYPOTHETICAL 50.0 KD P	1.40e+01

ALIGNMENTS

RESULT	1	PRELIMINARY;	PRT;	195 AA.
ID	Q63829			
AC	Q63829;			
DT	01-NOV-1996 (TrEMBLrel. 01, Created)			
DT	01-NOV-1996 (TrEMBLrel. 01, Last sequence update)			
DT	01-NOV-1999 (TrEMBLrel. 12, Last annotation update)			
DE	BUP-5'OF BMI-1 PROVIRAL INSERTION LOCUS.			
GN	BUP.			
OS	Mus musculus (Mouse).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;			
OC	Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE; 93165034.			
RA	HAUPT Y., BARRI G., ADAMS J.M.;			
RT	"Nucleotide sequence of bup, an upstream gene in the bmi-1 proviral			
RT	insertion locus."			
RL	Mol. Biol. Rep. 17:17-20(1992).			
DR	EMBL; S54914; AAB25378.1; -.			
DR	MGD; MGI:88218; Bup.			
SQ	SEQUENCE 195 AA; 22037 MW; 827A84F1 CRC32;			

Query Match	8.0%;	Score 124;	DB 11;	Length 195;
Best Local Similarity	22.1%;	Pred. No. 4.75e-03;		
Matches	21;	Conservative 33;	Mismatches 37;	Indels 4; Gaps 4;
Db	98	ELFCTEQNNKNSLETGLSRLPHITDVSRLEYQIKTNQLHKMYRPGYLYLVNEN	157	
QY	125	DLASLAFSGRPLDLSVAQQQSSLPVSYFRVRVDAISTSAQSRSLQPSVLMLQLKLT	184	
Db	158	NDSQSYEINFSCNWEQLQDLVG-KLKDASKSLER	191	
QY	185	GSAHRF-EVPIA-KFQELRYSVALLVKEMAE-LEK	216	
RESULT	2	PRELIMINARY;	PRT;	491 AA.
ID	P73944			
AC	P73944;			
DT	01-FEB-1997 (TrEMBLrel. 02, Created)			
DT	01-FEB-1997 (TrEMBLrel. 02, Last sequence update)			
DT	01-JAN-1999 (TrEMBLrel. 09, Last annotation update)			
DE	HYPOTHETICAL 55.6 KD PROTEIN.			
OS	Synechocystis sp. (strain PCC 6803).			
OC	Bacteria; Cyanobacteria; Chroococcales; Synechocystis.			

ID	Q43642	PRELIMINARY;	PR	PT	649	AA.
AC	Q43642;					
DT	01-JUN-1998	(TrEMBLrel. 06, Created)				
DT	01-JUN-1998	(TrEMBLrel. 06, Last sequence update)				
DT	01-NOV-1999	(TrEMBLrel. 12, Last annotation update)				
DE	HEAT SHOCK PROTEIN 75.					
GN	HSP75.					
OS	Homo sapiens (Human).					
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;					
OC	Eutheria; Primates; Catarrhini; Hominidae; Homo.					
RN	[1]					
RN	SEQUENCE FROM N.A.					
RP	MEDLINE; 96347537.					
RP	CHEN C.F., CHEN Y., DAI K., CHEN P.L., RILEY D.J., LEE W.H.;					
RL	Submitted (JAN-1998) to the EMBL/GenBank/DBJ databases.					
RL	EMBL; AF043254; AAC02679.1; -					
DR	HSP; P02829; 1A8.					
DR	PFAM; PF00183; HSP90; 2.					
DR	PRINTS; PR00775; HEATSHOCK90.					
DR	Heat shock.					
KW						
SEQUENCE	649	AA; 74017	MW; CCAF787C	CRC32;		
Query Match	7.7%	Score 120;	DB 4;	Length 649;		
Best Local Similarity	22.3%	Pred. No. 1.54e-02;				
Matches	23;	Conservative 40;	Mismatches 34;	Indels 6;	Gaps 5;	
Db	43	KKLLDIVARSLXS-E-KEYFIRELISNADALEKRLKLVSGQALPEMHIQTNAERG 100				
QY	49	RKLLKLVGALHGKDCREA-VEQLGASANLSERLAV-LLAGTHTLLOQALRLPPASLKP 106				
Db	101	TTITQD-TGIGTQEEVLNIGTIARSGSKAFDLALONQAEAS 142				
QY	107	DARQEELEQLGIPQD-LIGDLASLAFSGQRPLDSVAQOQGS 148				
RESULT	5					
ID	Q9ZWR95	PRELIMINARY;	PR	PT	512	AA.
AC	Q9ZWR95;					
DT	01-MAY-1999	(TrEMBLrel. 10, Created)				
DT	01-MAY-1999	(TrEMBLrel. 10, Last sequence update)				
DT	01-MAY-1999	(TrEMBLrel. 10, Last annotation update)				
DE	F5A8.3 PROTEIN.					
GN	F5A8.3.					
OS	Arabidopsis thaliana (Mouse-ear cress).					
OC	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;					
OC	euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons;					
OC	core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae;					
OC	Arabidopsis.					
RN	[1]					
RN	SEQUENCE FROM N.A.					
RC	STRAIN-CV. COLUMBIA.					
RA	FEDERSPEL N.A., PALM C.J., CONWAY A.B., KURTZ D.B., CONWAY A.R.,					
RA	AU M.H., ARAUJO R., BUEHLER E., DEWAR K., FENG J., KIM C., LI Y.,					
RA	QJI O., OSSORNE B.I., SHINN P., SUN H., TORIUMI M., VYZOSKAIA V.S.,					
RA	YU G., ECKER J., THEOLOGIS A., DAVIS R.W.;					
RL	Submitted (JAN-1999) to the EMBL/GenBank/DBJ databases.					
DR	EMBL; AC004146; AAD10659.1; -					
SEQUENCE	512	AA; 58463	MW; 9C771B39	CRC32;		
Query Match	7.6%	Score 118;	DB 10;	Length 512;		
Best Local Similarity	32.1%	Pred. No. 2.74e-02;				
Matches	25;	Conservative 20;	Mismatches 28;	Indels 5;	Gaps 5;	
Db	333	WTLLMLAHNPWTQDNNVRDVRVCGDGVPSVEQLSSL-TSLNKVINESLRLYPATLLP 391				
QY	49	RKLLKLVGALHGKDCREAVEQI-GASANLSERLAVLGGTHTLLQALRL-PASLKP 106				

```
Db 392 RMAFEDIKGLDIIIPKGL 409
QY 107 D-AFQE-ELQELGIPQDL 122

RESULT 6
ID Q27512 PRELIMINARY; PRT; 676 AA.
AC Q27512;
DT 01-NOV-1996 (TReMBLrel. 01, Created)
DT 01-NOV-1996 (TReMBLrel. 01, Last sequence update)
DT 01-NOV-1999 (TReMBLrel. 12, Last annotation update)
DE NEX-2 PROTEIN.
GN NEX-2.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditidae;
OC Rhabditina; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
RN [1]
RP SEQUENCE FROM N.A.
RA BERKS M.;
RL Submitted (JAN-1994) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE; 94150718.
RA WILSON R., AINSCOUGH R., ANDERSON K., BAYNES C., BERKS M.,
RA BONFIELD J., BURTON J., CONNELL M., COPSEY T., COOPER J., COULSON A.,
RA CRAXTON M., DEAR S., DU Z., DURBIN R., FAVELLO A., FULTON L.,
RA GARDNER A., GREEN P., HAWKINS T., HILLIER L., JIER M., JOHNSTON L.,
RA JONES M., KERSHAW J., KIRSTEN J., LAISTER N., LATREILLE P.,
RA LIGHTNING J., LLOYD C., MCMURRAY A., MORTIMORE B., O'CALLAGHAN M.,
RA PARSONS J., PERCY C., RIFKEN L., ROOPRA A., SAUNDERS D., SHOWNKEEN R.,
RA SWALDON N., SMITH A., SONNHAMMER E., STADEN R., SULSTON J.,
RA THIERRY-MIEG J., THOMAS K., VAUDIN M., VAUGHAN K., WATERSTON R.,
RA WATSON A., WEINSTOCK L., WILKINSON-SPROAT J., WOHLDMAN P.;
RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT elegans.";
RL Nature 368:32-38(1994).
CC -1- FUNCTION: CALCIUM/PHOSPHOLIPID-BINDING PROTEIN WHICH PROMOTES
CC MEMBRANE FUSION AND IS INVOLVED IN EXOCYTOSIS.
CC -1- SIMILARITY: TO OTHER PROTEINS OF THE ANNEXIN FAMILY.
DR EMBL; 229443; CA82571.1; -.
DR HSPF; P79134; IAVC.
DR PROSITE; PS00223; ANNEXIN; 1.
DR PFAM; PF00191; annexin; 4.
DR PRINTS; PR00196; ANNEXIN.
KW Annexin; Calcium/phospholipid-binding; Repeat.
SQ SEQUENCE 676 AA; 74969 MW; 00E36C79 CRC32;

Query Match 7.6%; Score 118; DB 5; Length 676;
Best Local Similarity 25.7%; Pred. No. 2.74e-02;
Matches 37; Conservative 31; Mismatches 67; Indels 9; Gaps 8;

Db 59 C-ELVKESRHEMLSEELATLYDCATSTRLPNDFADDQEDLKNIRYGLFKACHAIN 117
QY 64 CREAVEQLGASANLSEERAVLL--AGHTYLLQALRLPPASLPDA-FQEELQE-LGI- 118

Db 118 STDFYQKSEDLK-DEEKVALRIEIVEQKLAESAKNEKDIAWNVIVVAKSLRPVIN 176
QY 119 PODLIGDLASLAFGSRPLDLSVAQQ-QSSLPHVSYERWRVDVAISTSAQSRSLQPSVL 177

Db 177 IQMPTVGGDTN-FEDDIISFAQFR 199
QY 178 MQLKLTGSAHRFEVPIAKFOELR 201

RESULT 7
ID P72747 PRELIMINARY; PRT; 706 AA.
AC P72747;
DT 01-FEB-1997 (TReMBLrel. 02, Created)
DT 01-FEB-1997 (TReMBLrel. 02, Last sequence update)
DT 01-NOV-1999 (TReMBLrel. 12, Last annotation update)
DE HYPOTHETICAL 79.8 KD PROTEIN.
OS Synchocystis sp. (strain FCC 6803).
OC Bacteria; Cyanobacteria; Chroococcales; Synchocystis.
```

```
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-PCC6803;
RA TABATA S.;
RL Submitted (JUN-1996) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-PCC6803;
RX MEDLINE; 97061201.
RA KANEKO T., SATO S., KOTANI H., TANAKA A., ASAMIZU E., NAKAMURA Y.,
RA HATAJIMA N., HIROSAWA M., SUGIURA M., SASAMOTO S., KIMURA T.,
RA MIYAGUCHI T., MATSUO A., MURAKI A., NAKAZAKI N., NARAO K., OKUMURA S.,
RA SHIMO S., TAKEUCHI C., WADA T., WATANABE A., YAMADA M., YASUDA M.,
RA TABATA S.;
RT "Sequence analysis of the genome of the unicellular cyanobacterium
RT Synchocystis sp. strain PCC6803. II. Sequence determination of the
RT entire genome and assignment of potential protein-coding regions.";
RL DNA Res. 3:109-136(1996).
DR EMBL; D90900; BAA16762.1; -.
DR PFAM; PF00563; DUF2; 1.
DR PFAM; PF00990; DUF9; 1.
DR PFAM; PF00498; FHA; 1.
KW Hypothetical protein.
SQ SEQUENCE 706 AA; 79831 MW; 706E0D1E CRC32;

Query Match 7.1%; Score 110; DB 2; Length 706;
Best Local Similarity 27.7%; Pred. No. 2.64e-01;
Matches 23; Conservative 29; Mismatches 26; Indels 5; Gaps 5;

Db 585 RRAIEQIATIGVNLSDDFGIGL-GTSLCLQQ-FKIPAVKIH-ESFIKLEQSPVNEATII 641
QY 65 REAVEQLGASANLSEERAVLLAGHTYLLQALRLPPASLPDAFQEELQELGIPQDLI 123

Db 642 TSIMTLGRKLGVRISEGVETQQ 664
QY 124 GDLSLAFG-SQRPILDLSVAQQ 145

RESULT 8
ID O62159 PRELIMINARY; PRT; 470 AA.
AC O62159;
DT 01-AUG-1998 (TReMBLrel. 07, Created)
DT 01-AUG-1998 (TReMBLrel. 07, Last sequence update)
DT 01-NOV-1999 (TReMBLrel. 12, Last annotation update)
DE F13G11.1 PROTEIN.
GN F13G11.1.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditidae;
OC Rhabditina; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
RN [1]
RP SEQUENCE FROM N.A.
RA AINSCOUGH R.;
RL Submitted (DEC-1996) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE; 94150718.
RA WILSON R., AINSCOUGH R., ANDERSON K., BAYNES C., BERKS M.,
RA BONFIELD J., BURTON J., CONNELL M., COPSEY T., COOPER J., COULSON A.,
RA CRAXTON M., DEAR S., DU Z., DURBIN R., FAVELLO A., FULTON L.,
RA GARDNER A., GREEN P., HAWKINS T., HILLIER L., JIER M., JOHNSTON L.,
RA JONES M., KERSHAW J., KIRSTEN J., LAISTER N., LATREILLE P.,
RA LIGHTNING J., LLOYD C., MCMURRAY A., MORTIMORE B., O'CALLAGHAN M.,
RA PARSONS J., PERCY C., RIFKEN L., ROOPRA A., SAUNDERS D., SHOWNKEEN R.,
RA SWALDON N., SMITH A., SONNHAMMER E., STADEN R., SULSTON J.,
RA THIERRY-MIEG J., THOMAS K., VAUDIN M., VAUGHAN K., WATERSTON R.,
RA WATSON A., WEINSTOCK L., WILKINSON-SPROAT J., WOHLDMAN P.;
RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT elegans.";
RL Nature 368:32-38(1994).
DR EMBL; 283317; CAB05899.1; -.
DR PFAM; PF00751; DM-domain; 1.
SQ SEQUENCE 470 AA; 51820 MW; CC968597 CRC32;
```


RA	XUE Q.,	"Studies on the tail region of the temperate coliphage 186 genome.";	
RT	Thesis (1993), University of Adelaide, Australia.		
NRN	[2]		
NRN	SEQUENCE FROM N.A.		
NRN	MEDLINE; 98371265.		
NRN	PORTELLI R., DODD I.B., XUE Q., EGAN J.B.;		
NRN	"The late-expressed region of the temperate coliphage 186 genome.";		
NRN	Virology 248:117-130(1998).		
NRN	[3]		
NRN	SEQUENCE FROM N.A.		
NRN	DODD I.B., EGAN J.B.;		
NRN	"Bacteriophage 186 complete genome.";		
NRN	Submitted (OCT-1995) to the EMBL/GenBank/DBJ databases.		
NRN	[4]		
NRN	SEQUENCE FROM N.A.		
NRN	XUE Q., EGAN J.B.;		
NRN	Submitted (MAY-1998) to the EMBL/GenBank/DBJ databases.		
NRN	EMBL; U32222; AAC34162.1; -. S -> *		
DR	VARIANT 33 33		
FT	SEQUENCE 302 AA; 33005 MW; 5A210700 CRC32;		
SQ			
	Query Match	6.7%; Score 104; DB 9; Length 302;	
	Best Local Similarity	25.2%; Pred. No. 1.35e+00;	
	Matches	37; Conservative 37; Mismatches 67; Indels 6; Gaps 6;	
Db	135 MSVAGSTGAEFFH-GRSADRVDADISVISPSACVTISVLSRENNGAASDELSQSVRNAL 193		
QY	1 MSALGAAPYLLHPADSHGRVSLFGS-QPSPEVTVAQQLKDLDRSTFRKLLKLVYVAL 59		
Db	194 NAEQVRPVADRTVQSAQIIDYQIRATLTFVYFPGNEPIR-AAAEAKLKAYISAQHRLGR 252		
QY	60 HGKDCREAVEQUGA-SANLSEERLAVLAGTHTLLQQLRLPPASLKPDATQEEELQELGI 118		
Db	253 DTRLSAIIYAALHVEGVORVELAAPVAD 279		
QY	119 PODLIGDLASLAF-GSOR-PLLDLSVAQ 143		
RESULT	14		
ID	Q9XY72	PRELIMINARY; PRT; 382 AA.	
AC	Q9XY72;		
DT	01-NOV-1999 (TrEMBLrel. 12, Created)		
DT	01-NOV-1999 (TrEMBLrel. 12, Last sequence update)		
DT	01-NOV-1999 (TrEMBLrel. 12, Last annotation update)		
DE	PROBABLE ACYL-COA DEHYDROGENASE.		
GN	SC6A5.36.		
OS	Streptomyces coelicolor.		
OC	Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;		
OC	Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.		
RP	[1]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN-A3(2);		
RC	OLIVER K., HARRIS D.;		
RA	"A set of ordered cosmids and a detailed genetic and physical map for		
RA	the 8 Mb Streptomyces coelicolor A3(2) chromosome.";		
RL	Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.		
RL	[2]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN-A3(2);		
RA	BENTLEY S.D.; PARKHILL J., BARRELL B.G., RAJANDREAM M.A.;		
RL	Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.		
RP	[3]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN-A3(2);		
RC	STRAIN-A3(2);		
RX	MEDLINE; 97000351.		
RA	REDENBACH M., KIESER H.M., DENAPAITE D., EICHNER A., CULLUM J.,		
RA	KINASHI H., HOPWOOD D.A.;		
RT	"A set of ordered cosmids and a detailed genetic and physical map for		

RT the 8 Mb Streptomyces coelicolor A3(2) chromosome." ;
RL Mol. Microbiol. 21:77-96(1996).
DR EMBL; AL049485; CAB39720.1; -.
DR PROSITE; PS00072; ACYL_COA_DH.1; 1.
DR PROSITE; PS00073; ACYL_COA_DH.2; 1.
SQ SEQUENCE 382 AA; 42143 MW; 4E2F94FE CRC32;

Query Match
Best Local Similarity 6.7%; Score 104; DB 2; Length 382;
Matches 28; Conservative 27; Mismatches 41; Indels 7; Gaps 7;


Db 240 GNLQERLAIV-GAAATEEILDITTRYVKEAFGRPLAKLQHVFEIAEMATEA-AV 297
QY 75 ANLSERLAVLQTHLLQALRPASLK-PDAFQELQELGIPQDLIGDLASLAFGS 133
Db 298 TRTFDLRCVTEHVAGRLDPVHMAKWA-TQLQKRVVDCIQ 339
QY 134 QRPLDS-VAQQGSSL-P-HVSFRRVVDVAISTSAQSRSLQ 173

RESULT 15
ID 070211 PRELIMINARY; PRT; 603 AA.
AC 070211;
DT 01-AUG-1998 (TREMBlrel. 07, Created)
DT 01-AUG-1998 (TREMBlrel. 07, Last sequence update)
DT 01-NOV-1999 (TREMBlrel. 12, Last annotation update)
DE INSULIN-LIKE GROWTH FACTOR BINDING PROTEIN COMPLEX ACID-LABILE
DE SUBUNIT.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-FISHER;
EX MEDLINE; 98121980.
RA DELHANTY P.J., BAXTER R.C.;
RT "Cloning and characterization of the rat gene for the acid-labile
RT subunit of the insulin-like growth factor binding protein complex."
RL J. Mol. Endocrinol. 19:267-277(1997).
DR EMBL; AF006203; AAC15252.1; -.
DR HSP; P23945; IXUN.
DR PFAM; PF00580; LRR; 19.
DR PFAM; PF01463; LRRCT; 1.
DR PFAM; PF01462; LRRNT; 1.
DR PRINTS; PR00019; LEURICHRPT.
SQ SEQUENCE 603 AA; 66924 MW; 74E63165 CRC32;

Query Match
Best Local Similarity 6.7%; Score 104; DB 11; Length 603;
Matches 38; Conservative 31; Mismatches 36; Indels 14; Gaps 13;

Db 272 DLGSHNRVAGLMEDTFGLGLHLVLAHNAIASLRPTFKDLHFEELQIGHNRITQGE 331
QY 76 NLSEERLAVLAGT-HTLQ-QALRP-P-ASLKPDAFQE-B-LQELGIPQDLIGDLAS 128
Db 332 RTEGLGQLEVL-TLNDNQITEV-RVGAFLGNVAMNLSGNLSLPERVFGLDKL 388
QY 129 LAF-G-SORPLDSVAQQGSSLPHVSFRRVVDVAI-STSAQS-RSLQPSVLMQL-KL 182

Search completed: Fri May 12 10:25:41 2000
Job time : 87 secs.



Release 3.1A John F. Collins, BioComputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MPsrch_pp protein - protein database search, using Smith-Waterman algorithm
Run on: Fri May 12 10:23:17 2000; MasPar time 34.78 Seconds
192.339 Million cell updates/sec
Tabular output not generated.

Title: >US-09-223-796-2
Description: (1-224) from US09223796.ppe
Perfect Score: 1552
Sequence: 1 MSALGAAAPYLHPADSHSG.....ALVLKEMAELEKCKERKIQD 224

Scoring table: PAM 150
Gap 11

Searched: 82229 seqs, 29864866 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: swiss-prot38
1:swissprot

Statistics: Mean 47.784; Variance 95.999; scale 0.498

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
1	120	7.7	661	1	TRAL_HUMAN	3.38e-03
2	114	7.3	302	1	VPJ_BPP2	2.14e-02
3	112	7.2	677	1	BS4_MOUSE	3.92e-02
4	106	6.8	1170	1	SMC2_YEAST	2.30e-01
5	104	6.7	603	1	ALS_RAT	4.08e-01
6	104	6.7	886	1	YFIQ_ECOLI	5.43e-01
7	103	6.6	262	1	NCAP_INSVN	7.20e-01
8	102	6.6	556	1	HSCC_ECOLI	5.43e-01
9	103	6.6	566	1	YBAE_ECOLI	5.43e-01
10	102	6.6	603	1	ALS_MOUSE	7.20e-01
11	102	6.6	605	1	ALS_PAPPA	7.20e-01
12	100	6.4	219	1	YM50_MYCTU	1.26e+00
13	100	6.4	804	1	YBBP_ECOLI	1.26e+00
14	99	6.4	842	1	PKL2_RAT	1.66e+00
15	100	6.4	2156	1	RRPL_PUMH	1.26e+00
16	98	6.3	605	1	ALS_HUMAN	2.18e+00
17	98	6.3	2670	1	YAQS_SCHPO	2.18e+00
18	97	6.2	201	1	MGF_CHICK	2.87e+00
19	97	6.2	584	1	YMB3_YEAST	2.87e+00
20	97	6.2	1201	1	MEF_MYXXA	2.87e+00
21	96	6.2	1230	1	SMC3_YEAST	3.75e+00
22	95	6.1	63	1	YRKD_BACSU	4.91e+00
23	94	6.1	92	1	VNST_BUNSH	6.40e+00

RESULT	ID	TRAL_HUMAN	STANDARD;	PRT;	661 AA.
AC	Q12931	075235;			
DT	15-JUL-1999	(Rel. 38, Created)			
DT	15-JUL-1999	(Rel. 38, Last sequence update)			
DT	15-JUL-1999	(Rel. 38, Last annotation update)			
DE	TUMOR NECROSIS FACTOR TYPE 1 RECEPTOR ASSOCIATED PROTEIN (TRAP-1)				
DE	(FRAGMENT).				
GN	TRAP1.				
OS	Homo sapiens (Human).				
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;				
OC	Eutheria; Primates; Catarrhini; Hominidae; Homo.				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RX	MEDLINE; 95181307.				
RA	SONG H.Y., DUNBAR J.D., ZHANG Y.X., GUO D., DONNER D.B.;				
RT	"Identification of a protein with homology to hsp90 that binds the				
RT	type 1 tumor necrosis factor receptor.";				
RL	J. Biol. Chem. 270:3574-3581(1995).				
RN	[2]				
RP	SEQUENCE OF 16-631 FROM N.A.				
RA	RICKE D.O., BRUCE D., MUNDT M., DOGGETT N., MUNK C., SAUNDERS E.,				
RA	ROBINSON D., JONES M., BUCKINGHAM J., CHASTEN L., THOMPSON S.,				
RA	GOODWIN L., BRYANT J., TESMER J., MEINCKE L., LONGMIRE J., WHITE S.,				
RA	UENG S., TATUM O., CAMPBELL C., FAWCETT J., MALTBIE M., MISRA M.,				
RA	DAVERN L.;				
RL	Submitted (JUN-1998) to the EMBL/GenBank/DBJ databases.				
CC	-!- FUNCTION: BINDS TO THE INTRACELLULAR DOMAIN OF TUMOR NECROSIS				
CC	FACTOR TYPE 1 RECEPTOR.				
CC	-!- TISSUE SPECIFICITY: FOUND IN SKELETAL MUSCLE, LIVER, HEART, BRAIN,				
CC	KIDNEY, PANCREAS, LUNG AND PLACENTA.				
CC	-!- SIMILARITY: BELONGS TO THE HEAT SHOCK PROTEIN 90 FAMILY.				
CC	-----				
CC	THIS SWISS-PROT entry is copyright. It is produced through a collaboration				
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -				
CC	the European Bioinformatics Institute. There are no restrictions on its				
CC	use by non-profit institutions as long as its content is in no way				
CC	modified and this statement is not removed. Usage by and for commercial				
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/				
CC	or send an email to license@isb-sib.ch).				
CC	-----				
CC	EMBL; U12595; AAC87704.1; -				
DR	EMBL; AC005203; AAC24722.1; -				
DR	PROSITE; PS00298; HSP90; FALSE_NEG.				
DR	PFAM; PF00183; HSP90; 2.				
FT	NON_TER	1			
FT	CONFLICT	292	292		
				G -> R (IN REF. 2).	

```
SQ SEQUENCE 661 AA; 75342 MW; 055E95AF CRC32;
Query Match 7.7%; Score 120; DB 1; Length 661;
Best Local Similarity 22.3%; Pred. No. 3.38e-03;
Matches 23; Conservative 40; Mismatches 34; Indels 6; Gaps 6;

Db 80 KKLDDIVARSYS-E-KVFTRILSNASDALEKLRHKLKLVSDGQALPMEIHLQTNAEKG 137
QY 49 RKLKLVGVALHGKDCREA-VEQLGASANLSEERLAV-LLAGTHTLLQQLRPPASLKP 106
Db 138 TITQD-TGICMTOBELVSLGTARGSKAFDALQNAEAS 179
QY 107 DAFQELQELIPQD-LIGDLASLAFGSRPLDSVAQQGSS 148

RESULT 2
ID VFJ_BPP2 STANDARD; PRT; 302 AA.
AC P51767;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 15-DEC-1999 (Rel. 39, Last annotation update)
DE BASEPLATE ASSEMBLY PROTEIN J (GPU).
GN J.
OS Bacteriophage P2.
OS Viruses; dsDNA viruses, no RNA stage; Tailed phages; Myoviridae.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 96036485.
RA HAGGARD-LJUNGQUIST E., JACOBSEN E., RISHOVD S., SIX E.W., NILSEN O.,
RA SUNSHINE M.G., LINDQVIST B.H., KIM K.-J., BARREIRO V., KOONIN E.V.,
RA CALENDAR R.
RT "Bacteriophage P2: genes involved in baseplate assembly.";
RL Virology 213:109-121(1995).
CC -1- FUNCTION: LIES AT THE EDGE OF THE BASEPLATE.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AF063097; AAD03284.1; -
SQ SEQUENCE 302 AA; 32780 MW; CB8BA88 CRC32;
Query Match 7.3%; Score 114; DB 1; Length 302;
Best Local Similarity 25.9%; Pred. No. 2.14e-02;
Matches 37; Conservative 35; Mismatches 66; Indels 5; Gaps 5;

Db 135 LSVAGSVGAYQYH-GRSADGRVADISVTPSPACVTSVLSRENNGVASEDLLAVVRNAL 193
QY 1 MSALGAAPYLHPADSHSGRVSLG-SQPSPEVTAVALKLDLDRSTFRKLLKLVVVAL 59
Db 194 NGEDVRFVADVTVQSAIVYQINATLYLPPESEPIR-AAVKKLEAVITQAHLGR 252
QY 60 HGKDCRAVEVLGA-SANLSEERLAVLAGHTHTLLQQLRPPASLKPDAFQELQELGI 118
Db 253 DIRLSAIYAALHVEGVQVELAA 275
QY 119 PQDLIGDLASLAF-GSORPLDS 140

RESULT 3
ID B54_MOUSE STANDARD; PRT; 677 AA.
AC P54729;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE B54 PROTEIN.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 95212908.
RA STRUNNIKOV A.V., HOGAN E., KOSHLAND D.;
RT "SMC2, a Saccharomyces cerevisiae gene essential for chromosome
RT segregation and condensation, defines a subgroup within the SMC
RT family.";
RL Genes Dev. 9:587-599(1995).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-S288C / AB972;
RX MEDLINE; 95400292.
RA MURAKAMI Y., NAITOU M., HAGIWARA H., SHIBATA T., OZAWA M.,
RA SASANUMA S.-I., SASANUMA M., TSUCHIYA Y., SOEDA E., YOKOYAMA K.,
RA YAMAZAKI M., TASHIRO H., EKI T.;
RT "Analysis of the nucleotide sequence of chromosome VI from
RT Saccharomyces cerevisiae.";
RL Nat. Genet. 10:261-268(1995).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN-S288C / AB972;
RX MEDLINE; 96287654.
RA EKI T., NAITOU M., HAGIWARA H., ABE M., OZAWA M., SASANUMA S.-I.,
RA SASANUMA M., TSUCHIYA Y., SHIBATA T., WTANABE K., ONO A.,
RA YAMAZAKI M.-A., TASHIRO H., HANAOKA F., MURAKAMI Y.;
RT "Fifteen open reading frames in a 30.8 kb region of the right arm of
```



```
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DE HYPOTHETICAL 98.0 KD PROTEIN IN UNG-PSSA INTERGENIC REGION.
GN YFIQ.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-K12 / MG1655;
RX MEDLINE: 97426617.
RA BLATTNER F.R., PLUNKETT G. III, BLOCH C.A., PERNA N.T., BURLAND V.,
RA RILEY M., COLLADO-VIDES J., GLASNER F.D., RODE C.K., MAYHEW G.F.,
RA GREGOR J., DAVIS N.W., KIRKPATRICK H.A., GOEDEN M.A., ROSE D.J.,
RA MAU B., SHAO Y.;
RT "The complete genome sequence of Escherichia coli K-12.";
RL Science 277:1453-1474(1997).
RN [2]
RP SEQUENCE OF 1-612 FROM N.A.
RC STRAIN-K12;
RA NASHIMOTO H., SAITO N.;
RL Submitted (SEP-1995) to the EMBL/GenBank/DBJ databases.
CC -!- SIMILARITY: SOME, IN THE N-TERMINAL, TO SUCCINYL-COA LIGASES.
CC -!- SIMILARITY: STRONG, TO M.JANNASCHII M70590.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announcement/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: A5000344; AAC75637.1; -.
CC EMBL: D64044; BAA10925.1; -.
CC EMBL: D64044; YFIQ.
CC ECGENE: EGI4224; YFIQ.
CC PFAM: PF00593; Acetyltransf; 1.
CC Hypothetical protein.
CC CONFLICT 506 507 ST -> YA (IN REF. 2).
CC SEQUENCE 886 AA; 97987 MW; 5D400142 CRC32;
Query Match 6.78; Score 104; DB 1; Length 886;
Best Local Similarity 27.28; Pred. No. 4.08e-01;
Matches 22; Conservative 19; Mismatches 39; Indels 1; Gaps 1;
Db 416 RLSEAGLPTRTPEGTITAFMHVYRNKOLRETPALPSN-LTSNTAEHLQQAI 474
QY 38 QLKLDLRSIFRKLKLVGALHGKDCREAVEQLGASNLSEERLAVLLAGTHLQQAL 97
Db 475 AEGATSLDTHVEVQPIQYGM 495
QY 98 RLPPASLKPDFAFOELQELGI 118
RESULT 7
ID NCAP_INSVN STANDARD; PRT; 262 AA.
AC Q01808;
DT 01-JUL-1993 (Rel. 26, Created)
DT 01-JUL-1993 (Rel. 26, Last sequence update)
DT 01-FEB-1994 (Rel. 28, Last annotation update)
DE NUCLEOCAPSID PROTEIN (NUCLEOPROTEIN).
GN N.
OS Impatiens necrotic spot virus (strain NL-07) (INSV).
OC Viruses; ssRNA negative-strand viruses; Bunyaviridae; Tospovirus.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE: 92331780.
RA DE HAAN P., DE AVILA A.C., KORMELINK R., WESTERBROEK A.,
RA GELLEN J.J.B., PETERS D., GOLDBACH R.;
RT "The nucleotide sequence of the S RNA of Impatiens necrotic spot
RT virus, a novel tospovirus.";
RL FEBS Lett. 306:27-32(1992).
CC -----
CC -!- FUNCTION: ENCAPSIDATES THE RNA.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announcement/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: X66972; CAA47383.1; -.
CC PIR: S23159; S23159.
CC PFAM: PF01533; Tospo_nucleocap; 1.
CC Nucleocapsid; RNA-binding.
CC SEQUENCE 262 AA; 28718 MW; 1B526B0F CRC32;
Query Match 6.68; Score 103; DB 1; Length 262;
Best Local Similarity 36.24; Pred. No. 5.43e-01;
Matches 17; Conservative 12; Mismatches 17; Indels 1; Gaps 1;
Db 137 ATTDKSCVLGVLGGSLPIASVLFETAAALVPAIYQDAKHVELGI 183
QY 73 ASANLSEERLAVLLAGTHLQQALRLPASLKPDAFOELQELGI 118
RESULT 8
ID HSCC_ECOLI STANDARD; PRT; 556 AA.
AC P77319;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 15-DEC-1999 (Rel. 39, Last annotation update)
DE CHAPERONE PROTEIN HSCC (HSC62).
GN HSCC.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-K12 / MG1655;
RX MEDLINE: 97426617.
RA BLATTNER F.R., PLUNKETT G. III, BLOCH C.A., PERNA N.T., BURLAND V.,
RA RILEY M., COLLADO-VIDES J., GLASNER F.D., RODE C.K., MAYHEW G.F.,
RA GREGOR J., DAVIS N.W., KIRKPATRICK H.A., GOEDEN M.A., ROSE D.J.,
RA MAU B., SHAO Y.;
RT "The complete genome sequence of Escherichia coli K-12.";
RL Science 277:1453-1474(1997).
RN [2]
RP SEQUENCE FROM N.A.
RA CHUNG E., ALLEN E., ARAUJO R., APARICIO A., DAVIS K., DUNCAN M.,
RA FEDERSPIEL N., HYMAN R., KALMAN S., KOMP C., KURDI O., LEW H., LIN D.,
RA NAMATH A., OEFNER P., ROBERTS D., SCHRAMM S., DAVIS R.W.;
RL Submitted (JAN-1997) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN-K12;
RX MEDLINE: 97061202.
RA OSHIMA T., AIBA H., BABA T., FUJITA K., HAYASHI K., HONJO A.,
RA IKEMOTO K., INADA T., ITOH T., KAJIHARA M., KANAI K., KASHIMOTO K.,
RA KIMURA S., KITAGAWA M., MAKINO K., MASUDA S., MIKI T., MIYBUCHI K.,
RA MORI H., MOTOMURA K., NAKAMURA Y., NASHIMOTO H., NISHIO Y., SAITO N.,
RA SAMPEI G., SEKI Y., TAGAMI H., TAKEMOTO K., WADA C., YAMAMOTO Y.,
RA YANO M., HORIUCHI T.;
RT "A 718-Kb DNA sequence of the Escherichia coli K-12 genome
RT corresponding to the 12.7-28.0 min region on the linkage map.";
RL DNA Res. 3:137-155(1996).
RN [4]
RP CHARACTERIZATION.
RX MEDLINE: 98407901.
RA YOSHIMUNE K., YOSHIMURA T., ESAKI N.;
RT "Hsc62, a new DnaK homologue of Escherichia coli.";
RL Biochem. Biophys. Res. Commun. 250:115-118(1998).
CC -!- FUNCTION: PROBABLE CHAPERONE. HAS ATPASE ACTIVITY. NOT STIMULATED
CC BY DNAJ.
```

```
CC --!- SIMILARITY: BELONGS TO THE HEAT SHOCK PROTEIN 70 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch)
CC -----
DR EMBL; AE000169; AAC73751.1; -
DR EMBL; U82598; ABA40851.1; -
DR EMBL; D90705; BAA35302.1; -
DR EMBL; D90704; BAA35297.1; -
DR HSSP; P19120; IATS;
DR ECGENE; EG13653; HSCC;
DR PROSITE; PS00297; HSP70_1; 1;
DR PROSITE; PS00329; HSP70_2; 1;
DR PROSITE; PS01036; HSP70_3; 1;
DR PFAM; PF00012; HSP70; 1;
KW Chaperone; ATP-binding.
SQ SEQUENCE 556 AA; 61986 MW; 44794D92 CRC32;
Query Match 6.6%; Score 102; DB 1; Length 556;
Best Local Similarity 23.1%; Pred. No. 7.20e-01;
Matches 24; Conservative 33; Mismatches 43; Indels 4; Gaps 4;
Db 269 EDLWPLNLRV-PIEQALDARKPSQIDSLVGGASOMPLVQRIAVRLFGLPYQS 327
QY :||| : : : : : : : : : : : : : : : : : : : : : : : : :
95 QALRPP-ASLKPDAFQELQELGIPQDLIGLASLAFGSQRPLDLSVAQQGSSLPVHS 153
: ||| : : : : : : : : : : : : : : : : : : : : : : : : :
Db 328 YDPSTI-VALGAAIQACRLSRSEIEVILTDICPYSLGVEYNR 370
YD :||| : : : : : : : : : : : : : : : : : : : : : : : : :
QY 154 YFRVRVDVAISTSAQSR-SLOPSVLMLKLTGDSAHREVPPIAK 196
: ||| : : : : : : : : : : : : : : : : : : : : : : : : :
RESULT 9 STANDARD; PRT; 566 AA.
ID YBAE_ECOLI
AC P46890; P77158;
DC 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DE INSULIN-LIKE GROWTH FACTOR BINDING PROTEIN COMPLEX ACID LABILE CHAIN
DE PRECURSOR (ALS).
DE HYPOTHETICAL 65.0 KD PROTEIN IN HUPB-COF INTERGENIC REGION.
GN YBAE.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-K12 / MG1655;
RX MEDLINE; 97426617.
RA BLATTNER F.R., PLUNKETT G. III, BLOCH C.A., PERNA N.T., BURLAND V.,
RA RILEY M., COLLADO-VIDES J., GLASNER F.D., RODE C.K., MAYHEW G.F.,
RA GREGOR J., DAVIS N.W., KIRKPATRICK H.A., GOEDEN M.A., ROSE D.J.,
RA MAU B., SHAO Y.;
RT "The complete genome sequence of Escherichia coli K-12."
RL Science 277:1453-1474(1997).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-K12 / MG1655;
RA ROBERTS D., ALLEN E., ARAUTO R., APARICIO A., CHUNG E., DAVIS K.,
RA DUNCAN M., FEDERSPIEL N., HYMAN R., KALMAN S., KOMP C., KURDI O.,
RA LEW H., LIN D., NAMATH A., OEFNER P., SCHRAMM S., DAVIS R.W.;
RL Submitted (JAN-1997) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN-K12 / W3110;
RA HAYADA E., OHMORI H., QIAO Y., TSUJI M., FUKUDA R.;
RL Submitted (OCT-1996) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE OF 1-358 FROM N.A.
RC STRAIN-K12 / MC4100;
RA PATZER S.I., HANTKE K.;
```

```
RL Submitted (OCT-1995) to the EMBL/GenBank/DBJ databases.
CC --!- SIMILARITY: TO E. COLI YABN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch)
CC -----
DR EMBL; AE000151; AAC73548.1; -
DR EMBL; U82664; ABA40201.1; -
DR EMBL; D82943; BAA11649.1; -
DR EMBL; Z54355; CAA91182.1; -
DR ECGENE; EG13217; YBAE.
DR PFAM; PF00496; SHP_bac_5; 2.
KW Hypothetical protein.
SQ SEQUENCE 566 AA; 64991 MW; 091FA6F1 CRC32;
Query Match 6.6%; Score 103; DB 1; Length 566;
Best Local Similarity 35.4%; Pred. No. 5.43e-01;
Matches 23; Conservative 16; Mismatches 19; Indels 7; Gaps 7;
Db 67 LRFVLPESLRNMMEQAL-ETGKQDVL-ELAQAPGELRLLQPFMGQWQNDPTLR 124
QY :||| : : : : : : : : : : : : : : : : : : : : : : : : :
97 LR-L-PPASLKPDAFQELQELGIPQDLIGLASLAFGSQRPLDLS-VA-Q-QQGSSLP 151
: ||| : : : : : : : : : : : : : : : : : : : : : : : : :
Db 125 IPYR 129
YD :||| : : : : : : : : : : : : : : : : : : : : : : : : :
QY 152 VSYER 156
: ||| : : : : : : : : : : : : : : : : : : : : : : : : :
RESULT 10 STANDARD; PRT; 603 AA.
ID ALS_MOUSE
AC P70389;
DC 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DE INSULIN-LIKE GROWTH FACTOR BINDING PROTEIN COMPLEX ACID LABILE CHAIN
DE PRECURSOR (ALS).
DE IGFALS OR ALS OR ALBS.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-129/SV;
RX MEDLINE; 96413591.
RA BOISCLAIR Y.R., SETO D., HSIEH S., HURST K.R., OOI G.T.;
RT "Organization and chromosomal localization of the gene encoding the
RT mouse acid labile subunit of the insulin-like growth factor binding
RT complex."
RL Proc. Natl. Acad. Sci. U.S.A. 93:10028-10033(1996).
CC --!- FUNCTION: MAY HAVE AN IMPORTANT ROLE IN REGULATING THE ACCESS OF
CC CIRCULATING IGFs TO THE TISSUES.
CC --!- SUBUNIT: FORMS A TERNARY COMPLEX OF ABOUT 140 TO 150 KD WITH IGF-I
CC OR IGF-II AND IGFBP-3 (BY SIMILARITY).
CC --!- SUBCELLULAR LOCATION: EXTRACELLULAR.
CC --!- SIMILARITY: THE REPEATED LEUCINE-RICH (LRR) SEGMENT IS FOUND IN
CC MANY PROTEINS. NUMBER IN THIS PROTEIN: 20.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch)
CC -----
DR EMBL; U66900; AAB17270.1; -
DR MGD; MGI:107973; IGFALS.
DR PFAM; PF00560; LRR; 19.
```


Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.

[1]
SEQUENCE FROM N.A.
RC STRAIN=H37RV;
RX MEDLINE: 98295987.
RA COLE S.T., BROSCH R., PARKHILL J., GARNIER T., CHURCHER C., HARRIS D.,
GORDON S.V., EGLMEIER K., GAS S., BARRY C.E. III, TEKAIA F.,
RA BADCOCK K., BASHAM D., BROWN D., CHILLINGWORTH T., CONNOR R.,
RA DAVIES K., DEVLIN K., FELTWELL T., GENTLES S., HAMLIN N., HOLROYD S.,
RA HORNSBY T., JAGELS K., KROGH A., MCLEAN J., MOULE S., MURPHY L.,
RA OLIVER S., OSBORNE J., QUAIL M.A., RAJANDREAM M.A., ROGERS J.,
RA RUTTER S., SEEGER K., SKELTON S., SQUARES S., SQUARES R., SULSTON J.E.,
RA TAYLOR K., WHITEHEAD S., BARRELL B.G.;
RT "Deciphering the biology of Mycobacterium tuberculosis from the
NT complete genome sequence";
RL Nature 393:537-544(1998).
CC -|- SIMILARITY: BELONGS TO THE TETR/ACRR FAMILY OF TRANSCRIPTIONAL
CC REGULATORS.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).

DR EMBL; Z70692; CAA94669.1; .
DR PROSITE; PS01081; HTH_TETR_FAMILY; FALSE_NEG.
DR PFAM; PF00440; tetr; 1.
DR KW Hypothetical protein; Transcription regulation; DNA-binding.
FT DNAStrand 62 81 H-T-H MOTIF (BY SIMILARITY).
KW DNA_BIND
SQ SEQUENCE 219 AA; 23875 MW; F3F30AD6 CRC32;

Query Match 6.4%; Score 100; DB 1; Length 219;
Best Local Similarity 25.6%; Pred. No. 1.26e+00;
Matches 23; Conservative 20; Mismatches 42; Indels 5; Gaps 5;

Dbb 54 VDYGVDVTLAEIIRACGVSPTV-YRQWPDTSTRSMASMLTSHI-ADVLRVPLDGDDRE 111
Yy 129 LAFGQRPLDSVAQQQGSSLPHYSFWRVDVAISTAQSRSLQP SVLMQLKLTGSAH 188
Db 112 ALVKIOIVAVADR LR-GDDLIMSVMHSELAR 140
Yy 189 RFEPVPIAFQE-LRYSVVALVKEM-AELEK 216

RESULT 13
ID YBBP_ECOLI STANDARD; PRT; 804 AA.
AC PF77504;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE HYPOTHETICAL 89.3 KD PROTEIN IN TESA-RHSD INTERGENIC REGION.
GN YBBP.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
[1]
RP SEQUENCE FROM N.A.
RC STRAIN-K12 / MG1655;
RX MEDLINE: 97426617.
RA BLATTNER F.R., PLUNKETT G. III, BLOCH C.A., PERNA N.T., BURLAND V.,
RA RILEY M., COLLADO-VIDES J., GLASNER F.D., RODE C.K., MAYHEW G.F.,
RA GREGOR J., DAVID N.W., KIRKPATRICK H.A., GOEDEN M.A., ROSE D.J.,
RA MAU B., SHAO Y.;
RT "The complete genome sequence of Escherichia coli K-12.";
RL Science 277:1453-1474(1997).
[2]
RP SEQUENCE FROM N.A.
RA ROBERTS D., ALLEN E., ARAUTO R., APARICIO A., CHUNG E., DAVIS K.,
RA DUNCAN M., FEDERSPIEL N., HYMAN R., KALMAN S., KOMP C., KURDI O.,

(TM)

Result No.	Query %			DB	ID	Description	Pred. No.
	Score	Match	Length				
1	121	7.8	491	2	S75449	hypothetical protein	1.52e+02
2	120	7.7	661	2	A35877	tumor necrosis factor	2.01e+02
3	118	7.6	676	2	S41022	hypothetical protein	3.51e+02
4	110	7.1	706	2	T74610	hypothetical protein	3.12e+01
5	107	6.9	668	2	T11778	phosphoglycerate tran	6.92e+01
6	107	6.9	913	2	T15474	hypothetical protein	6.92e+01
7	106	6.8	1170	2	A36157	chromosome segregatio	8.99e+01
8	104	6.7	228	2	T72540	hypothetical protein	1.51e+00
9	104	6.7	603	3	JG1282	insulin-like growth f	1.51e+00
10	104	6.7	886	2	G65036	hypothetical protein	1.51e+00
11	104	6.7	1116	2	S77213	DNA-directed DNA poly	1.51e+00
12	102	6.6	159	2	T05656	hypothetical protein	2.52e+00
13	103	6.6	262	2	S23159	nucleoprotein N - Imp	1.96e+00
14	102	6.6	451	2	S74728	hypothetical protein	2.52e+00
15	103	6.6	555	2	F72111	fructose-6-p phosphot	1.96e+00
16	102	6.6	556	2	H64799	probable dnaK-type mo	2.52e+00
17	103	6.6	566	2	E64774	probable membrane pro	1.96e+00
18	102	6.6	603	2	JG6128	insulin-like growth f	2.52e+00
19	102	6.6	605	2	JG5239	insulin-like growth f	2.52e+00
20	100	6.4	219	2	F70779	hypothetical protein	4.19e+00
21	100	6.4	362	2	D72734	probable histidine de	4.19e+00
22	99	6.4	371	2	T56863	hypothetical protein	5.38e+00
23	100	6.4	804	2	G64780	probable membrane pro	4.19e+00

[illegible]

```
##molecule_type DNA
##residues 1-1170 ##label STR
##cross-references GB:U05820; NID:g468039; PIDN:AAA17416.1; PID:g468040
REFERENCE S56186
#authors Murakami, Y.; Naitou, M.; Hagiwara, H.; Shibata, T.; Ozawa, M.; Sasanuma, S.I.; Sasanuma, M.; Tsuchiya, Y.; Soeda, E.;
```

Haikawa, Y.; Jin-no, K.; Takahashi, M.; Sekine, M.; Baba, S.; Ankai, A.; Kosugi, H.; Hosoyama, A.; Fukui, S.; Nagai, Y.; Nishijima, K.; Nakazawa, H.; Takamiya, M.; Masuda, S.; Funahashi, T.; Tanaka, T.; Kudoh, Y.; Yamazaki, J.; Kushida, N.; Oguchi, A.; Aoki, K.; Kubota, K.; Nakamura, Y.; Nomura, N.; Sako, Y.; Kikuchi, H.

#journal	DNA Res. (1999) 6:83-101
#title	Complete genome sequence of an aerobic hyper-thermophilic Crenarchaeon, Aeropyrum pernix K1.
#cross-references	MUID:99310339
#accession	E72540
##status	preliminary
##molecule_type	DNA
##residues	1-228 ##label KAW
##cross-references	DBJ:AP000062; NID:g5105244; PID:BAA80610.1;
##experimental_source	PID:g1044396; PID:g5105297
GENETICS	
##gene	
SUMMARY	
APEL10	
length 228	#molecular-weight 25890 #checksum 7192
Query Match	6.7%; Score 104; DB 2; Length 228;
Best Local Similarity	32.2%; Pred. No. 1.51e+00;
Matches	29; Conservative 22; Mismatches 33; Indels 6; Gaps 6;
Db	131 VVQLVEKDVGSPRRRLKRLKPDLAAYATGGDALAVARELWG-RRLARLGAVFRVSI 189
Qy	36 VAQLT-KDLDKSTFR-KLLKLGVALLGKDCRAVEOLGASNLSEERLAVLLAGTH-TL 92
Db	190 VPSRRLVAVASPRPDSFGPWWEEVLISIVQ 219
Qy	93 LQ-QALRLPPASLKPDAFQELQE-LGIPQ 120
RESULT	9
ENTRY	JCI282
TITLE	insulin-like growth factor-binding protein acid labile chain
ORGANISM	precursor - rat
DATE	30-Sep-1993 #sequence_revision 30-Sep-1993 #text_change 15-Aug-1997
ACCESSIONS	JCI282
REFERENCE	JCI282
#authors	Dal, J.; Baxter, R.C.
#journal	Biochem. Biophys. Res. Commun. (1992) 188:304-309
#title	Molecular cloning of the acid-labile subunit of the rat insulin-like growth factor binding protein complex.
#cross-references	MUID:93038676
#accession	JCI282
##molecule_type	mRNA
##residues	1-603 ##label DAI
##experimental_source	liver
##note	the authors translated the codon AAG for residue 63 as Arg, AAA for residue 205 as Pro and GGT for residue 260 as Arg
CLASSIFICATION	#superfamily leucine-rich alpha-2-glycoprotein repeat homology
FEATURE	
1-27	#domain signal sequence #status predicted #label SIG\
28-603	#product insulin-like growth factor binding protein, acid labile chain #status predicted #label MAT
SUMMARY	length 603 #molecular-weight 66811 #checksum 8075
Query Match	6.7%; Score 104; DB 2; Length 603;
Best Local Similarity	31.9%; Pred. No. 1.51e+00;
Matches	38; Conservative 31; Mismatches 36; Indels 14; Gaps 13;
Db	272 DLSNRRVAGLWEDTFPCLLGLHLVRLAHNAIAISLRPTFKDLHFLEQLGHNRIQLGE 331
Qy	76 NLSEERLAVLLAGT-HTLLQ-QALRLP--P-ASLKPDAPQE-E-LQELGIPQDLGLDAS 128
Db	332 RTFEGLGQLEVL-TLNDNQITEV-RVGAFGSLGNVAMNLSGNCRLSLSPERFQGLDKL 388
Qy	129 LAF-G-SQRPILDSVAQQGSSLPVYSFRWRVDVAI-STSAQS-RSLQPSVLMLQ-LK 182
RESULT	10
ENTRY	G65036
TITLE	hypothetical protein b2584 - Escherichia coli (strain K-12)

```

Query Match      6.6%; Score 103; DB 2; Length 262;
Best Local Similarity 36.2%; Pred. No. 1.96e+00;
Matches 17; Conservative 12; Mismatches 17; Indels 1; Gaps 1;

db      137 ATTTDLKSCVLGVLLGGSGUPLTASVNFETIAALVPAYIQDAKHVELGI 183
QV      73 ASANISEERLAVLLAGTHTTLLQQALRLPPASLKPDAFQELQ-ELGI 118

```

RESULT	14
ENTRY	S74728
TITLE	#type complete hypothetical protein slr0971 ccma 3'-region - <i>Synechocystis</i> sp. (strain PCC 6803)
ORGANISM	#formal_name <i>Synechocystis</i> sp. PCC 6803
#variety	25-Apr-1997 #sequence_revision 25-Apr-1997 #text_change 21-Aug-1998
DATE	S74728; B55214
ACCESSIONS	S74322
REFERENCE	Kaneko, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asamizu, E.; Nakamura, Y.; Miyajima, N.; Hirose, M.; Sugita, M.; Kamura, T.; Hosouchi, T.; Matsuno, A.; Sasamoto, S.
#authors	

```

S.; Iaredunir, C.; Roda, I.; Macanase, A.; Temada, A. 1977
Yasuda, M.; Tabata, S.
#journal      DNA Res. (1996) 3:109-136
#title        Sequence analysis of the genome of the unicellular
               cyanobacterium Synechocystis sp. PCC6803. II. Sequence
               determination of the entire genome and assignment of
               potential protein-coding regions.
#cross-references MIMD:197061201
#accession    S74728
#status        nucleic acid sequence not shown; translation not shown
#molecule_type DNA
#residues      1-451 #label KAN
#cross-references EMBL:D50901; GB:AB001339; NID:g1651897; PID:d1017612
               PID:g1651953
#note          the nucleotide sequence was submitted to the EMBL Data
               Library, June 1996

```

```

#authors J. Bacteriol. (1994) 176:2374-2378
#journal
#title A gene (ccma) required for carboxysome formation in the
cyanobacterium Synechocystis sp. strain PCC6803.
#cross-references MUID:94209239
#accession B5214
##molecule_type DNA
##residues
'VDPNLEFVAQISRLMQLAGGPGLLFENVK',
'SPEFVAVNLGCTVERICWAMNMHPLELDLGGK',244-451
OXA
##cross-references GB:D26444; NID:g452237; PID:d1006006; PID:g499299
##note the authors translated the codon CCC for residue 236 as
Glu; the authors translated used GTG as an initiation
codon and translated it as Val
##note the authors say that codon usages suggest this protein
is not expressed at high levels
#length 451 #molecular_weight 49667 #checksum 7763
SUMMARY

```

Query Match	6.68;	Score 102;	DB 2;	Length 451;
Best Local Similarity	45.58;	Pred. No. 2.52e+00;		
Matches	25;	Conservative	5;	Mismatches 21; Indels 4; Gaps 4;
Db	42	ALHFDPRPPEVQQLTSGT-PEE-LEALALAVETLYQORLN-PCASLSPSTYPE	93	
Ov	58	ALHGKDCR-EAVELGASNLSEERLAVLIAGHTHTLLOALRPASLPKDAFQE	111	

RESULT	15
ENTRY	F72111
TITLE	#type complete fructose-6-p phosphotransferase - Chlamydia pneumoniae (strain CWL023)
ORGANISM	#formal name Chlamydia pneumoniae

```

##residues      1-262  ##label  HAA
##cross-references EMBL:X66972; NID:g60488; PTDN:CAA47383.1; PID:g60490
#superfamily tomato spotted wilt virus nucleoprotein N
#length 262 #molecular-weight 28718 #checksum 7578
CLASSIFICATION
SUMMARY

```

```

DATE          23-Apr-1999 #sequence_revision 23-Apr-1999 #text_change
              20-Sep-1999
- ACCESSIONS F72111
REFERENCE     A72000
#authors      Kalman, S.; Mitchell, W.; Marathe, R.; Lammel, C.; Fan, J.;
              Olinger, L.; Grimwood, J.; Davis, R.W.; Stephens, R.S.
#journal      Nature Genet. (1999) 21:385-389
#title        Comparative genomes of Chlamydia pneumoniae and C.
              trachomatis.
#cross-references MUID:99206606
#accession     F72111
              #status      preliminary
              #molecule_type DNA
              ##residues    1-555 #label ARN
              ##cross-references GB:AE001603; GB:AE001363; NID:g4376422; PID:g4376428
              ##experimental_source strain CWL029
GENETICS
#gene          pfkA_1
CLASSIFICATION #superfamily pyrophosphate--fructose-6-phosphate
              1-phosphotransferase alpha chain; 6-phosphofructokinase 1
              homology
SUMMARY        #length 555 #molecular-weight 62414 #checksum 6064
Query Match    6.68; Score 103; DB 2; Length 555;
Best Local Similarity 23.1%; Pred. No. 1.96e+00;
Matches 21; Conservative 29; Mismatches 35; Indels 6; Gaps 6;
Db 315 IEHFDTRKLIDE-LNVLGANGDSIEKILSKLPETLTFHLPKDIANQLLARDSHG 373
QY 68 VEQLGASANLSEERLAVLLAGTHTLLQQAL-RDPPASLKP-DAFQBEL-OELGIPQDLIG 124
Db 374 NRVYSKIATE-ELLAVMVKKEIKRPHMEF 403
QY 125 DLASLAFGSRPLLDVAQQQSSSL-PHVSYS 154

Search completed: Fri May 12 10:26:57 2000
Job time : 58 secs.

```

[M][P][S][R][L] (TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MPsrch_pp protein - protein database search, using Smith-Waterman algorithm
Run on: Fri May 12 10:28:07 2000; MasPar time 17.96 Seconds
Tabular output not generated. 161.600 Million cell updates/sec

Title: >US-09-223-796-2
Description: (1-224) from US09223796.pep
Perfect Score: 1552
Sequence: 1 MSALGAAPYLHHPADSHG.....ALVLKEMAELEKKCKERKLD 224

Scoring table: PAM 150
Gap 11

Searched: 131253 seqs, 12956647 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: a-issued
1:5A_COMB 2:5B_COMB 3:PCT_COMB 4:backfiles1

Statistics: Mean 30.351; Variance 153.535; scale 0.198

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
1	124	8.0	195	2	US-08-822-Sequence 3, Applicatio	2.71e+01
2	118	7.6	195	2	US-08-822-Sequence 1, Applicatio	7.00e+01
3	104	6.7	603	1	US-08-190-Sequence 50, Applicati	6.12e+00
4	98	6.3	605	1	US-08-190-Sequence 49, Applicati	1.51e+01
5	95	6.1	912	1	US-07-789-Sequence 8, Applicatio	2.37e+01
6	95	6.1	912	1	US-08-487-Sequence 8, Applicatio	2.37e+01
7	95	6.1	912	1	US-08-005-Sequence 8, Applicatio	2.37e+01
8	94	6.1	4472	2	US-08-804-Sequence 2, Applicatio	2.75e+01
9	93	6.0	405	3	PCT-US93-1-Sequence 2, Applicatio	3.18e+01
10	92	5.9	681	2	US-08-346-Sequence 37, Applicati	3.69e+01
11	92	5.9	681	1	US-08-083-Sequence 16, Applicati	3.69e+01
12	92	5.9	1078	1	US-08-083-Sequence 11, Applicati	3.69e+01
13	92	5.9	1078	2	US-08-346-Sequence 32, Applicati	3.69e+01
14	92	5.9	1078	1	US-08-465-Sequence 32, Applicati	3.69e+01
15	92	5.9	1078	1	US-08-264-Sequence 32, Applicati	3.69e+01
16	92	5.9	2544	2	US-08-576-Sequence 20, Applicati	3.69e+01
17	92	5.9	2556	1	US-08-083-Sequence 32, Applicati	3.69e+01
18	88	5.7	301	2	US-08-420-Sequence 47, Applicati	6.61e+01
19	88	5.7	301	2	US-08-343-Sequence 22, Applicati	6.61e+01
20	89	5.7	1114	2	US-08-576-Sequence 31, Applicati	5.71e+01
21	87	5.6	122	2	US-08-475-Sequence 19, Applicati	7.63e+01
22	87	5.6	122	1	US-08-476-Sequence 19, Applicati	7.63e+01
23	87	5.6	122	1	US-07-956-Sequence 19, Applicati	7.63e+01

24	87	5.6	122	1	US-08-485-Sequence 19, Applicati	7.63e+01
25	87	5.6	2396	1	US-08-157-Sequence 2, Applicatio	7.63e+01
26	85	5.5	365	2	US-08-481-Sequence 10, Applicati	1.02e+02
27	86	5.5	368	1	US-08-423-Sequence 35, Applicati	8.81e+01
28	85	5.5	379	1	US-08-484-Sequence 2, Applicatio	1.02e+02
29	85	5.5	379	1	US-08-347-Sequence 2, Applicatio	1.02e+02
30	85	5.5	379	3	PCT-US95-0-Sequence 4, Applicatio	1.02e+02
31	85	5.5	379	3	PCT-US94-0-Sequence 2, Applicatio	1.02e+02
32	85	5.5	379	3	PCT-US95-1-Sequence 2, Applicatio	1.02e+02
33	85	5.5	379	3	PCT-US95-1-Sequence 2, Applicatio	1.02e+02
34	86	5.5	393	1	US-08-423-Sequence 33, Applicati	8.81e+01
35	86	5.5	393	1	US-08-530-Sequence 11, Applicati	8.81e+01
36	85	5.5	560	1	US-07-683-Sequence 1, Applicatio	1.02e+02
37	86	5.5	746	3	PCT-US95-1-Sequence 2, Applicatio	8.81e+01
38	86	5.5	2254	2	US-08-677-Sequence 3, Applicatio	8.81e+01
39	86	5.5	2254	2	US-08-790-Sequence 3, Applicatio	8.81e+01
40	86	5.5	2482	1	US-08-328-Sequence 6, Applicatio	8.81e+01
41	85	5.5	2509	1	US-08-469-Sequence 10, Applicati	1.02e+02
42	85	5.5	2556	1	US-08-185-Sequence 17, Applicati	1.02e+02
43	86	5.5	3248	1	US-08-353-Sequence 1, Applicatio	8.81e+01
44	86	5.5	3248	3	PCT-US95-1-Sequence 1, Applicatio	8.81e+01
45	85	5.5	15281	2	US-08-471-Sequence 2, Applicatio	1.02e+02

ALIGNMENTS

RESULT 1
ID US-08-822-260-3 STANDARD; PRT; 195 AA.
XX xxxxxx
XX
DT
XX
DE Sequence 3, Application US/08822260
XX
CC Sequence 3, Application US/08822260
CC Patent No. 5830660
CC GENERAL INFORMATION:
CC APPLICANT: Hillman, Jennifer L.
CC APPLICANT: Goli, Surya K.
CC TITLE OF INVENTION: NOVEL TUMORIGENESIS PROTEIN
CC NUMBER OF SEQUENCES: 3
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: Incyte Pharmaceuticals, Inc.
CC STREET: 3174 Porter Drive
CC CITY: Palo Alto
CC STATE: CA
CC COUNTRY: USA
CC ZIP: 94304
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Diskette
CC COMPUTER: IBM Compatible
CC OPERATING SYSTEM: DOS
CC SOFTWARE: FastSeq for Windows Version 2.0
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/08/822,260
CC FILING DATE: Herewith
CC CLASSIFICATION: 514
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER:
CC FILING DATE:
CC ATTORNEY/AGENT INFORMATION:
CC NAME: Billings, Lucy J.
CC REGISTRATION NUMBER: 36,749
CC REFERENCE/DOCKET NUMBER: PF-0247 US
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: 415-855-0555
CC TELEFAX: 415-845-4166
CC TELEX:
CC INFORMATION FOR SEQ ID NO: 3:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 195 amino acids
CC TYPE: amino acid

Db 272 DLSHNRVAGLMEDTFPGLGLHVLRLAHNAIASLRPTFKDLHLEELQIGHNRIRQLGE 331

CC SOFTWARE: PatentIn Release #1.0, Version #1.30

CC CURRENT APPLICATION DATA:

CC APPLICATION NUMBER: US/08/465,500

CC FILING DATE: 05-JUN-1995

CC CLASSIFICATION: 435

CC ATTORNEY/AGENT INFORMATION:

CC NAME: Mistoek, S. Leslie

CC REGISTRATION NUMBER: 18,872

CC REFERENCE/DOCKET NUMBER: 7326-034

CC TELECOMMUNICATION INFORMATION:

CC TELEPHONE: (212) 790-9090

CC TELEFAX: (212) 869-8864/9741

CC INFORMATION FOR SEQ ID NO: 32:

CC SEQUENCE CHARACTERISTICS:

CC LENGTH: 1078 amino acids

CC TYPE: amino acid

CC STRANDEDNESS: single

CC TOPOLOGY: unknown

CC MOLECULE TYPE: protein

SQ SEQUENCE 1078 AA; 116254 MW; 5971068 CN;

Query Match 5.9%; Score 92; DB 1; Length 1078;
Best Local Similarity 29.9%; Pred.No. 3.69e+01;
Matches 20; Conservative 15; Mismatches 29; Indels 3; Gaps 2;

D b 893 MVGPLHSSLAASALSMYSVQLPSTRLATPHLYVTQQVPQNLMQMQQNLPANIQQO 952
::|||:: ||:: |::||| :: |:: |:: |:: |:: |:: |:: |:: |:: |::
Q Y 55 VVGALHGKDCRAVEGLGASANLSERLAV--LAGTHILLOALLRPLPASLKPDFAFOEE 112

D b 953 -QSLOPP 958
||| | |
Q Y 113 LQELGP 119

RESULT 15

ID US-08-264-534-32 STANDARD; PRT; 1078 AA.

XX xxxxxx

AC

XX

DT

XX

DE Sequence 32, Application US/08264534

XX

Sequence 32, Application US/08264534

PATENT No. 5648464

GENERAL INFORMATION:

APPLICANT: Artavanis-Tsakonas, Spyridon et al.

TITLE OF INVENTION: Human No. 5648464ch And Delta, Binding Domains

TITLE OF INVENTION: In Toporythmic Proteins, And Methods Based Thereon

NUMBER OF SEQUENCES: 34

CORRESPONDENCE ADDRESS:

ADDRESSEE: Pennie & Edmonds

STREET: 1155 Avenue of the Americas

CITY: New York

STATE: New York

COUNTRY: U.S.A.

ZIP: 10036

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/264,534

FILING DATE:

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/695,189

FILING DATE: 03-MAY-1991

ATTORNEY/AGENT INFORMATION:

NAME: Mistoek, S. Leslie

REGISTRATION NUMBER: 18,872

REFERENCE/DOCKET NUMBER: 7326-004

Search completed: Fri May 12 10:28:29 2000
Job time : 22 secs.

MPSrch_pp protein - protein database search, using Smith-Waterman algorithm
Run on: Fri May 12 10:27:14 2000; MasPar time 29.86 Seconds
Tabular output not generated.
Title: >US-09-223-796-2
Description: (1-224) from US09223796.pep
Perfect Score: 1552
Sequence: 1 MSALGAAAPYLHPADSHG.....ALVLKEMAELEKCKERKLQD 224
Scoring table: PAM 150
Gap 11
Searched: 18963 seqs, 23686106 residues
Post-processing: Minimum Match 0%
Listing first 45 summaries
Database: a-geneseq36
l:geneseq
Statistics: Mean 32.859; Variance 162.439; scale 0.202
Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.
SUMMARIES
Result No. Score Match Length DB ID Description Pred. No.
1 1552 100.0 224 1 W37723 Rat Hypertension relat 2.90e-120
2 119 7.7 169 1 W88390 Human Zneul partial po 1.44e+00
3 119 7.7 181 1 W88391 Human Zneul partial po 1.44e+00
4 119 7.7 254 1 W88382 Human neuro-growth fac 1.44e+00
5 119 7.7 273 1 W88381 Human neuro-growth fac 1.44e+00
6 118 7.6 195 1 W71884 Amino acid sequence of 1.69e+00
7 118 7.6 196 1 Y07867 Human secreted protein 1.69e+00
8 104 6.7 603 1 R85889 WD-40 domain-contg. ra 1.46e+01
9 103 6.6 262 1 R42079 Impatiens Necrotic Spo 1.69e+01
10 98 6.3 605 1 R58888 WD-40 domain-contg. in 3.57e+01
11 97 6.2 178 1 W95014 Chicken MGF sequence. 4.14e+01
12 95 6.1 912 1 R36731 Ubiquitin-specific pro 5.56e+01
13 94 6.1 4472 1 W22601 Tyllactone synthase ORF 6.43e+01
14 93 6.0 166 1 W05049 Human thrombopoietin d 7.44e+01
15 93 6.0 405 1 R57140 Mouse mucosal addressin 7.44e+01
16 93 6.0 405 1 W60615 Mouse mucosal addressin 7.44e+01
17 93 6.0 547 1 R26189 Epidemic NANBH virus a 7.44e+01
18 91 5.9 193 1 W14362 Erythropoietin variant 9.95e+01
19 91 5.9 390 1 W37585 An AC1-1 polypeptide. 9.95e+01
20 91 5.9 920 1 W82500 Human hN3K protein. 9.95e+01
21 91 5.9 1078 1 R28963 Notch hN3K full length 9.95e+01
22 90 5.8 796 1 R84618 Oilseed rape acetyl Co 1.15e+02
23 89 5.7 332 1 R99525 Neutral protease of Cl 1.133e+02

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd
MPSrch_pp protein - protein database search, using Smith-Waterman algorithm
Run on: Fri May 12 10:27:14 2000; MasPar time 29.86 Seconds
Tabular output not generated.
Title: >US-09-223-796-2
Description: (1-224) from US09223796.pep
Perfect Score: 1552
Sequence: 1 MSALGAAAPYLHPADSHG.....ALVLKEMAELEKCKERKLQD 224
Scoring table: PAM 150
Gap 11
Searched: 18963 seqs, 23686106 residues
Post-processing: Minimum Match 0%
Listing first 45 summaries
Database: a-geneseq36
l:geneseq
Statistics: Mean 32.859; Variance 162.439; scale 0.202
Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.
SUMMARIES
Result No. Score Match Length DB ID Description Pred. No.
1 1552 100.0 224 1 W37723 Rat Hypertension relat 2.90e-120
2 119 7.7 169 1 W88390 Human Zneul partial po 1.44e+00
3 119 7.7 181 1 W88391 Human Zneul partial po 1.44e+00
4 119 7.7 254 1 W88382 Human neuro-growth fac 1.44e+00
5 119 7.7 273 1 W88381 Human neuro-growth fac 1.44e+00
6 118 7.6 195 1 W71884 Amino acid sequence of 1.69e+00
7 118 7.6 196 1 Y07867 Human secreted protein 1.69e+00
8 104 6.7 603 1 R85889 WD-40 domain-contg. ra 1.46e+01
9 103 6.6 262 1 R42079 Impatiens Necrotic Spo 1.69e+01
10 98 6.3 605 1 R58888 WD-40 domain-contg. in 3.57e+01
11 97 6.2 178 1 W95014 Chicken MGF sequence. 4.14e+01
12 95 6.1 912 1 R36731 Ubiquitin-specific pro 5.56e+01
13 94 6.1 4472 1 W22601 Tyllactone synthase ORF 6.43e+01
14 93 6.0 166 1 W05049 Human thrombopoietin d 7.44e+01
15 93 6.0 405 1 R57140 Mouse mucosal addressin 7.44e+01
16 93 6.0 405 1 W60615 Mouse mucosal addressin 7.44e+01
17 93 6.0 547 1 R26189 Epidemic NANBH virus a 7.44e+01
18 91 5.9 193 1 W14362 Erythropoietin variant 9.95e+01
19 91 5.9 390 1 W37585 An AC1-1 polypeptide. 9.95e+01
20 91 5.9 920 1 W82500 Human hN3K protein. 9.95e+01
21 91 5.9 1078 1 R28963 Notch hN3K full length 9.95e+01
22 90 5.8 796 1 R84618 Oilseed rape acetyl Co 1.15e+02
23 89 5.7 332 1 R99525 Neutral protease of Cl 1.133e+02

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd
MPSrch_pp protein - protein database search, using Smith-Waterman algorithm
Run on: Fri May 12 10:27:14 2000; MasPar time 29.86 Seconds
Tabular output not generated.
Title: >US-09-223-796-2
Description: (1-224) from US09223796.pep
Perfect Score: 1552
Sequence: 1 MSALGAAAPYLHPADSHG.....ALVLKEMAELEKCKERKLQD 224
Scoring table: PAM 150
Gap 11
Searched: 18963 seqs, 23686106 residues
Post-processing: Minimum Match 0%
Listing first 45 summaries
Database: a-geneseq36
l:geneseq
Statistics: Mean 32.859; Variance 162.439; scale 0.202
Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.
SUMMARIES
Result No. Score Match Length DB ID Description Pred. No.
1 1552 100.0 224 1 W37723 Rat Hypertension relat 2.90e-120
2 119 7.7 169 1 W88390 Human Zneul partial po 1.44e+00
3 119 7.7 181 1 W88391 Human Zneul partial po 1.44e+00
4 119 7.7 254 1 W88382 Human neuro-growth fac 1.44e+00
5 119 7.7 273 1 W88381 Human neuro-growth fac 1.44e+00
6 118 7.6 195 1 W71884 Amino acid sequence of 1.69e+00
7 118 7.6 196 1 Y07867 Human secreted protein 1.69e+00
8 104 6.7 603 1 R85889 WD-40 domain-contg. ra 1.46e+01
9 103 6.6 262 1 R42079 Impatiens Necrotic Spo 1.69e+01
10 98 6.3 605 1 R58888 WD-40 domain-contg. in 3.57e+01
11 97 6.2 178 1 W95014 Chicken MGF sequence. 4.14e+01
12 95 6.1 912 1 R36731 Ubiquitin-specific pro 5.56e+01
13 94 6.1 4472 1 W22601 Tyllactone synthase ORF 6.43e+01
14 93 6.0 166 1 W05049 Human thrombopoietin d 7.44e+01
15 93 6.0 405 1 R57140 Mouse mucosal addressin 7.44e+01
16 93 6.0 405 1 W60615 Mouse mucosal addressin 7.44e+01
17 93 6.0 547 1 R26189 Epidemic NANBH virus a 7.44e+01
18 91 5.9 193 1 W14362 Erythropoietin variant 9.95e+01
19 91 5.9 390 1 W37585 An AC1-1 polypeptide. 9.95e+01
20 91 5.9 920 1 W82500 Human hN3K protein. 9.95e+01
21 91 5.9 1078 1 R28963 Notch hN3K full length 9.95e+01
22 90 5.8 796 1 R84618 Oilseed rape acetyl Co 1.15e+02
23 89 5.7 332 1 R99525 Neutral protease of Cl 1.133e+02

ALIGNMENTS

RESULT 1
ID W37723 standard; Protein; 224 AA.
AC W37723;
DT 09-JUN-1998 (first entry)
DE Rat Hypertension related calcium regulated.
KW Hypertension related calcium regulated gene; HcARG; rat parathyroid;
KW extracellular calcium concentration; antibody; hypertension;
KW hyperthyroidism; osteoporosis; heart failure; diabetes; stroke;
KW cancer; inflammatory disease; asthma.
OS Rattus rattus.
FH Key Location/Qualifiers
FT Misc_difference 15..21 /note= "EF-hand like motif"
FT W09749807-A2.
PN 31-DEC-1997.
PD 23-JUN-1997; CA0439.
PR 21-JUN-1996; US-667495.
PA (GOSS/) GOSSARD F.
PA (HAME/) HAMEY P.
PA (LEWA/) LEWANCZUK R.
PA (TREMA/) TREMBLAY J.
PI Gossard F, Hamet P, Lewanczuk R, Tremblay J;
DR WPI; 98-077171/07.
DR N-PSDB; V18890.
PT Hypertension related calcium regulated gene - useful to develop
PT products to treat or detect, e.g. hypertension, stroke,
PT osteoporosis, heart failure, cancer, diabetes or asthma
PS Claim 8; Pages 26-27; 46pp; English.
CC This is the amino acid sequence of the hypertension related calcium
CC regulated gene (HcARG), which was isolated from the rat parathyroid.
CC Its expression is regulated by extracellular calcium concentration.
CC An antibody against the protein, can be used to detect or modulate
CC (e.g. enhance or inhibit) abnormal calcium levels. They can
CC specifically be used to detect or treat, e.g. hypertension,
CC hyperthyroidism, osteoporosis, heart failure, diabetes, stroke, cancer,
CC inflammatory disease, and asthma.
SQ Sequence 224 AA;

Query Match 100.0%; Score 1552; DB 1; Length 224;
Best Local Similarity 100.0%; Pred. No. 2.90e-120;
Matches 224; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Db 1 MSALGAAAPYLHPADSHGSRVFLGSPPEVTAVALKLDLDRSTFRKLLKLVGALH 60
|||||
Qy 1 MSALGAAAPYLHPADSHGSRVFLGSPPEVTAVALKLDLDRSTFRKLLKLVGALH 60
Db 61 GKDCRAVQLGASNLSEERLAVLLAGHTLLQQLRUPASLKPDAPFOELQELGIPQ 120

```

PR 18-JUN-1997; US-878322.
PR 18-JUN-1997; US-050143.
PA (ZYMO ) ZYMOGENETICS INC.
PI Blumberg H, Jelinek LJ, Lehner JM, Sheppard PO, Whitmore TE.
DR WPI; 99-095324/08.
FT New mammalian zneul polypeptides - used to, e.g. treat Alzheimer's
PS disease, cancer and to repopulate blood cells
PS Claim 6; Page 58; 70pp; English.
CC This claimed polypeptide comprises amino acid residues 1-85 and
CC 159-254 of the human zneul mature protein (see W88382), i.e. the
CC first HSMHC3W5A homology domain fused to the second HSMHC3W5A
CC homology domain of zneul. zneul is a new neuro-growth factor-like
CC protein that can be used as a growth, maintenance, or
CC differentiation factor in the spinal cord, heart, spleen, testis,
CC thyroid and lymph nodes. zneul may also play a role in breast
CC cancer, glioblastomas, and pituitary adenomas. zneul can be used
CC to treat Alzheimer's disease, cancer, to repopulate blood cells
CC after chemotherapy, to stimulate myofibroblast proliferation,
CC stimulate or inhibit growth factors made in the placenta, in
CC fertility and contraception, or to regenerate nerves. Claimed
CC zneul polypeptides (see also W88382-97), including specific domains
CC of zneul and epitope-bearing portions of zneul, can be used to
CC raise specific antibodies for use e.g. in diagnostic assays.
SQ Sequence 181 AA;

Query Match          7.78; Score 119; DB 1; Length 181;
Best Local Similarity 42.08; Pred. No. 1.44e+00;
Matches      21; Conservative 13; Mismatches 13; Indels 3; Gaps

Db    105 KEEVQRQSRVDLEERKLQVLAPLHSLASQALEHGLPDPGSLVHSFQQ 154
      :|::||:::||::||::||::||::||::||::||::||::||::||::||:
Qy    65 REAVEQLGASANLSERLAVLLAGTHTLLQAL-R-LP-PASLKPDATQE 111

RESULT 4
AC W88382 standard; Protein; 254 AA.
ID W88382;
DT DT 26-APR-1999 (first entry)
DE Human neuro-growth factor-like protein zneul mature polypeptide.
KW Zneul; neuro-growth factor-like protein; human; breast cancer;
KW glioblastoma; pituitary adenoma; Alzheimer's disease; therapy;
KW nerve regeneration; haematopoiesis; fertility; contraception;
KW antibody.
OS Homo sapiens.
FH Key Location/Qualifiers
FT Domain 1..85 "hydrophilic domain (HSM1), homologous to
FT FT /note= "epidermal growth factor-like domain 1"
FT Domain 86..116
FT FT /note= "epidermal growth factor-like domain 2"
FT Domain 117..158
FT FT /note= "epidermal growth factor-like domain 2"
FT Domain 159..254
FT FT /note= "domain HSM2 homologous to an HSMHC3W5A
FT FT domain"
FN WO9857983-A2.
PN 23-DEC-1998.
PF PD 18-JUN-1998; U12763.
PP PR 18-JUN-1997; US-878322.
PR 18-JUN-1997; US-050143.
PA (ZYMO ) ZYMOGENETICS INC.
PI Blumberg H, Jelinek LJ, Lehner JM, Sheppard PO, Whitmore TE.
DR WPI; 99-095324/08.
FT New mammalian zneul polypeptides - used to, e.g. treat Alzheimer's
PS disease, cancer and to repopulate blood cells
PS Claim 6; Page 48-49; 70pp; English.
CC This polypeptide comprises human zneul mature polypeptide. zneul
CC is a new neuro-growth factor-like protein (see also W88381). Its
CC closest human homologue is HSMHC3W5A, a gene in the H1A class III
CC region, which is contained in a cosmid which contains Notch 4.
CC zneul is also homologous to Notch 4 in its EGF-like domains and may

```

CC be involved in EGF receptor pathways. Zneul is widely expressed in
 CC adult tissues, with high expression in heart, placenta, spleen,
 CC testis, thyroid, spinal cord and lymph node. Zneul can be used as
 CC a growth, maintenance, or differentiation factor in the spinal
 CC cord, heart, spleen, testis, thyroid and lymph nodes. It may
 CC play a role in breast cancer, glioblastomas, and pituitary
 CC adenomas. Zneul may also be used to treat Alzheimer's disease,
 CC cancer, to repopulate blood cells after chemotherapy, to stimulate
 CC myofibroblast proliferation, stimulate or inhibit growth factors
 CC made in the placenta, in fertility and contraception, or to
 CC regenerate nerves. Claimed zneul polypeptides (see also W88382-97),
 CC including specific domains of zneul and epitope-bearing portions of
 CC zneul, can be used to raise specific antibodies for use e.g. in
 CC diagnostic assays.
 SQ Sequence 254 AA;

Query Match 7.7%; Score 119; DB 1; Length 254;
 Best Local Similarity 42.0%; Pred. No. 1.44e+00;

Matches 21; Conservative 13; Mismatches 13; Indels 3; Gaps 3;

Db 178 KEVORLQSRVDLLEKQLVLAPLHSLASQALEHGLPDGSLVHSFQ 227

QY 65 REAVEQLGASANLSEERLAVLLAGTHTLLQAL-R-LP-PASLKPDFAQE 111

RESULT 5

ID W88381 standard; Protein; 273 AA.

AC W88381;

DE 26-APR-1999 (first entry)

DE Human neuro-growth factor-like protein Zneul.

KW Zneul-1; neuro-growth factor-like protein; human; breast cancer;

KW glioblastoma; pituitary adenoma; Alzheimer's disease; therapy;

KW nerve regeneration; haematopoiesis; fertility; contraception;

KW antibody.

OS Homo sapiens.

FH Key Location/Qualifiers

FT Peptide 1..19 putative signal peptide

FT Protein 20..273

FT Domain 20..104

FT Domain 105..135

FT Domain 136..177

FT Domain 178..273

FT Domain 105..135

FT Domain 136..177

FT Domain 178..273

FT Domain 105..135

FT Domain 136..177

FT Domain 178..273

FT Domain 105..135

FT Domain 136..177

FT Domain 178..273

FT Domain 105..135

FT Domain 136..177

FT Domain 178..273

FT Domain 105..135

FT Domain 136..177

FT Domain 178..273

FT Domain 105..135

FT Domain 136..177

FT Domain 178..273

FT Domain 105..135

FT Domain 136..177

FT Domain 178..273

FT Domain 105..135

FT Domain 136..177

FT Domain 178..273

FT Domain 105..135

FT Domain 136..177

FT Domain 178..273

FT Domain 105..135

FT Domain 136..177

FT Domain 178..273

FT Domain 105..135

FT Domain 136..177

FT Domain 178..273

CC differentiation factor in the spinal cord, heart, spleen, testis,
 CC thyroid and lymph nodes. It may also play a role in breast cancer,
 CC glioblastomas, and pituitary adenomas. Zneul may be used to treat
 CC Alzheimer's disease, cancer, to repopulate blood cells after
 CC chemotherapy, to stimulate myofibroblast proliferation, stimulate
 CC or inhibit growth factors made in the placenta, in fertility and
 CC contraception, or to regenerate nerves. Claimed zneul
 CC polypeptides (see also W88382-97), including specific domains of
 CC zneul and epitope-bearing portions of zneul, can be used to raise
 CC specific antibodies for use e.g. in diagnostic assays.
 SQ Sequence 273 AA;

Query Match 7.7%; Score 119; DB 1; Length 273;
 Best Local Similarity 42.0%; Pred. No. 1.44e+00;

Matches 21; Conservative 13; Mismatches 13; Indels 3; Gaps 3;

Db 197 KEVORLQSRVDLLEKQLVLAPLHSLASQALEHGLPDGSLVHSFQ 246

QY 65 REAVEQLGASANLSEERLAVLLAGTHTLLQAL-R-LP-PASLKPDFAQE 111

RESULT 6

ID W71684 standard; Protein; 195 AA.

AC W71684;

DE 04-DEC-1998 (first entry)

DE Amino acid sequence of the human tumorigenesis associated protein.

KW Human; tumorigenesis associated protein; HTAP; transplantation;

KW tumour; Antagonist; cancer; inflammation; immunological disease;

KW antibody; probe; primer; PCR; amplification; hybridisation;

KW inhibition.

OS Homo sapiens.

PN WO9841635-A1.

PD 24-SEP-1998.

PF 20-MAR-1998; U06066.

PR 20-MAR-1997; US-822260.

PA (INCY-) INCYTE PHARM INC.

PI Goli SK, Hillman JL;

DR WPI; 98-521224/44.

DR N-PSDB; V58281.

PT New tumorigenesis-associated protein and related nucleic acid,

PT vectors, transformed cells - antibodies, agonists and antagonists,

PT for diagnosis, treatment and prevention of abnormal cellular

PT differentiation, particularly cancers and inflammation

PS Claim 1; Fig 1A-1B; 54pp; English.

CC This is the amino acid sequence of the human tumorigenesis

CC associated protein (HTAP) used in the method of the invention. HTAP,

CC is involved with cell proliferation and inflammation. It can be used

CC to stimulate cell proliferation (e.g. of cells intended for

CC transplantation in treatment of tumours or infections, or to prevent

CC genetic defects). Antagonists of HTAP are used to treat or prevent

CC a wide range of cancers (adenocarcinoma, melanoma, sarcoma, lymphoma,

CC leukaemia etc.), also inflammation where associated with infection or

CC immunological disease (e.g. asthma, cystic fibrosis, rheumatoid

CC arthritis). HTAP is also used to raise antibodies and to screen

CC libraries for specific-binding agents. The antibodies are used as for

CC diagnosis or monitoring of HTAP-related diseases (in usual

CC its natural sources. HTAP derived probes or primers, are used in

CC standard amplification or hybridisation tests to diagnose HTAP-related

CC diseases; to identify related sequences; for genomic mapping and for

CC screening for specific inhibitors.

SQ Sequence 195 AA;

Query Match 7.6%; Score 118; DB 1; Length 195;

Best Local Similarity 23.2%; Pred. No. 1.69e+00;

Matches 22; Conservative 29; Mismatches 40; Indels 4; Gaps 4;

Db 98 ELFCFTEYQNNKNSLEILLGSGISGRSLPHITDVSRLVEYQIKTNQLHRMYPAYLVLTSLVQN 157

QY 125 DLASLAFSGSQRLDLSVAQQGSSLPFYSFWRVDVAISIAQSRSLQPSVLMQKLTLD 184

Db 158 TDSYPPIPSISCSMEQLQDLVG-KLKQASKSLER 191

RESULT

Example 4; Page 44-45; 64pp; English.
INSV RNA was purified from systemically infected *Nicotiana rustica*

DK P-FUBP; Q41290.
 PT Ubiquitin-specific protease(s) - specifically cleave fusion
 PT proteins regardless of size, at C-terminus of ubiquitin portion
 PS Clam 13; Page 58-64; 78p; English.
 PS The sequences given in R35729-31 represent the ubiquitin-specific
 CC proteases UBp1, UBp2, and UBp3. These proteases specifically cleave
 CC the 120 kD fusion protein ubiquitin-methionine-beta-galactosidase
 CC (ub-met-beta-gal) (see also R36728). The proteases UBp1 and UBp2
 CC demonstrate activity both in vivo and in vitro, whereas UBp3 is only
 CC active in vivo. These proteases have been shown to specifically
 CC cleave ub-met-beta-gal at the C-terminus of the ubiquitin moiety.
 CC The proteases can be used to deubiquitinate fusion proteins in vivo,
 CC therefore prokaryotic cells having an expression vector one of these
 CC proteases can be further transformed with an expression vector
 CC encoding a ubiquitin fusion protein. Such cells will then produce
 CC a deubiquitinated product having a predetermined N-terminal amino acid
 CC residue.
 SQ Sequence 912 AA:

QY 26 GSQPSVTV-AVAQLKDL-DRSTFRLLKLV---VGALHGKDCREAVEQIGASANLSEE 80

QY 81 RLAVLAGTHLLQQALRLPPA-SL-KP--DAFQEELQ-ELGIPQDLIGLASLAFGSQR 135
Db 4406 ALAAAYCALAEDDILRR-DVI.PPFI.TETLAAAI.TBOCPN 4441

QY 136 PLLDSVAQQQSSLPVSYFRWRV-DVAISTSAQSR 171

RESULT	14
ID	W05049 standard; Protein; 166 AA.
AC	W05049;
DT	10-DEC-1996 (first entry)
DE	Human thrombopoietin deletion variant drpo-6.
KW	Human; Thrombopoietin; TPO; variant; deletion mutant; megakaryocyte; growth factor; platelet production; thrombocytopoiesis.
KS	Swedish

	Synthetic.			
OS	J08168386-A.			
PN	PD			
PP	02-JUL-1996			
PR	13-JUL-1995;	199299		
PS	19-OCT-1994;	JP-278657.		
PT	(SNOW) SNOW BRAND MILK PROD CO LTD.			
PA	WPI; 96-357243/36.			
DR	N-PSDB; T41906.			
DT	Thrombopoietin variants having megakaryocyte growth factor activity			
PT	- useful for promoting growth of megakaryocytes and platelets, also			
PT	as biochemical assay reagent			
FPS	Claim 13; Page 15; 21pp; Japanese.			
CPC	Four deletion variants of thrombopoietin (TPO), designated dnpo-5,			
CC	-6,-7 and -8, were engineered. All four variants contained the			
CC	first 131 residues from the N-terminal region of human TPO and all			
CC	four variants promoted formation and growth of human megakaryocytes.			
CC	Proteins comprising at least residues 1-131 of TPO are useful for			
CC	promoting platelet production. The present sequence is that of dtpo-			
CC	which has the same sequence as wild-type TPO for the first 131			
CC	amino acids, followed by residues 319-353 from the C-terminus of			
CC	TPO.			
SC	Sequence	166 AA;		
Query Match		Score 93;	DB 1:	Length 166;
Best Local Similarity		26.9%;	Pred. No.	7.44e+01;
Mechanisms		20.	Growth Factor	1.31

Matches	39; Conservative	38; Mismatches	56; Indels	12; Gaps	13;
Dbb	15	TARLILSSPAPACDLRVLSKLLRD-SHYLHSLSOCPEVHPLPTVPLPAVDVSLGEWK	73		
QY	19	SGRVSLGSQLSP-EVTAVAAQLLKDLDRSTFRKLLK-LVVGALHGKDCREAVE-QLGA-S	74		
Dbb	74	TQWEETKAQDIL-GAVTILLEGVMAARGOLGPTCLSSLLGQLSGQVRLLLGALQSL-LGT	131		
QY	75	ANLSEERLAVLLAGHTHLQQALRUPPASKLPDAFQEELQEL-GIPQDLIGDLASLAFGS	133		
Dbb	132	QLHPLLLPDPSAPTPTPTSPLLNTSY	156		
QY	134	Q-RPLL-D-SV-AQQQSSPLPHVSY	154		
RESULT 15					
ID	R57140 standard; Protein; 405 AA.				
AC	R57140;				
DT	21-FEB-1995 (first entry)				
DE	Mouse mucosal addressin cell adhesion molecule (MAdCAM-1).				
KW	Mucosal addressin cell adhesion molecule; MAdCAM-1; leukocyte;				
KW	endothelium; transmembrane protein; immunoglobulin.				
OS	Mus musculus.				
FF	Key				
FF	Location/Qualifiers				
peptide	1..21				
FT	/note= "signal peptide"				
FT	362..385				
region	/note= "transmembrane region"				
FT	WO9413312-A.				
PPN	23-JUN-1994.				
PD					

PF 23-NOV-1993; U11404.
 PR 15-DEC-1992; US-990866.
 PA (STRD) UNIV LELAND STANFORD JUNIOR.
 PI Briskin M, Butcher E;
 DR WPI: 94-234206/28.
 DR N-PSDB; Q69224.
 PT DNA encoding a new mucosal addressin - is used to produce prods.
 PT for diagnosis and therapy, e.g. modulating homing of leukocytes
 PT to endothelial cells
 PS Disclosure: Page 21; 30pp; English.
 CC This mucosal addressin (MAd) protein is characterized by having
 CC 3 Ig-like domains, a mucine-like region between the 2nd and 3rd
 CC domains, a transmembrane domain and a cytoplasmic domain, where the
 CC MAd is capable of binding to a leukocyte surface membrane protein.
 CC The MAd cell adhesion molecule (MAdCAM-1) protein encoded by this
 CC cDNA can be used for screening for molecules capable of binding to
 CC MAdCAM-1 or inhibiting binding of leukocytes to MAdCAM-1. The
 CC protein may be used to treat intestinal inflammatory diseases and
 CC autoimmune diseases, and especially intestinal carcinomas, etc.
 SQ Sequence 405 AA;

Query Match 6.0%; Score 93; DB 1; Length 405;
 Best Local Similarity 31.6%; Pred. No. 7.44e-01;
 Matches 18; Conservative 20; Mismatches 13; Indels 6; Gaps 5;

Db 168 EEIQEAGTPLFRMTQ-RWRLP-SLGTAPP-ALHCOVTMQLPKLV--LTHRKEIPV 219
 QY 139 DSVAQGGSLPHVSYFRWRVDVAISTAQSRSLQPSVLMQL-KLTDGSAHREVP 194

Search completed: Fri May 12 10:27:49 2000
 Job time : 35 secs.

